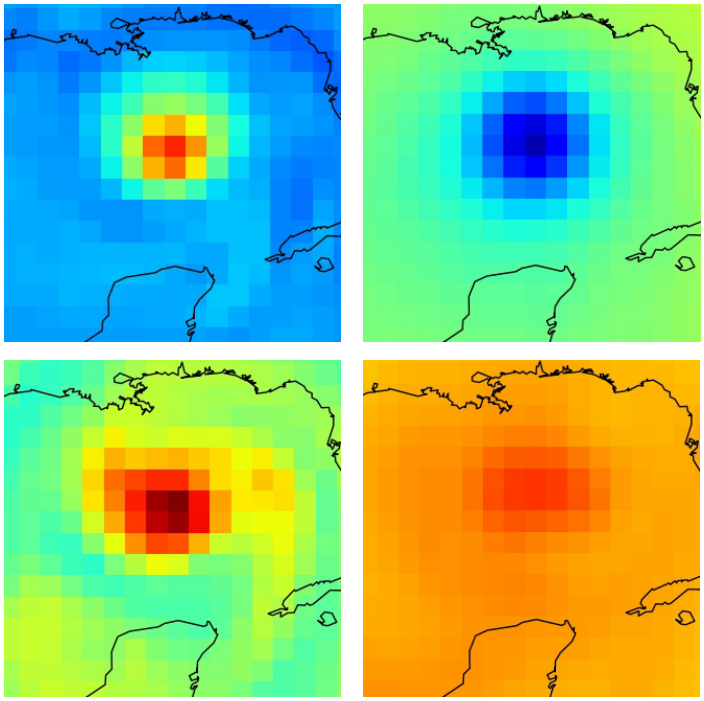
Detecting Activity of Tropical Cyclones with the Unsupervised Maximally Divergent Interval Algorithm

Detektion der Aktivität tropischer Zyklone mithilfe des unüberwachten Maximally Divergent Interval-Algorithmus

Master Thesis



Author: Simon Franz Zitzmann
Study program: Meteorology (M. Sc.)
Student ID: 11631020

Examiner: Prof. Dr. Markus Rapp
Scientific Supervisor: Prof. Dr. habil. Veronika Eyring

Faculty of Physics
Meteorological Institute
Ludwig Maximilian University of Munich

March 10, 2020

Abstract

The goal of this master thesis is the detection of tropical cyclones (TCs) by means of a multivariate unsupervised machine learning algorithm – the Maximally Divergent Intervals (MDI) method. Compared to traditional approaches of TC detection, it does not rely on hard, grid-size dependent thresholds of variables. Based on ERA-Interim reanalysis and the TC database IBTrACS, a labeled data set was created that allows the verification of the detections. The MDI algorithm is applied to the Gulf of Mexico during the hurricane seasons 2000 to 2010. First, the ideal initial settings were elaborated: this showed that the Kullback-Leibler divergence should be used to identify anomalies, no embedding should be applied and that the detections should be shifted one time step forward. It was subsequently found that the algorithm achieves the best detection skill with a mean Average Precision (mAP) of 0.537 when applied univariately to the relative vorticity at 850 hPa. A multivariate application involving other variables did not improve the mAP. In order to minimize the false alarm ratio, soft variable thresholds in wind speed of $8\,\mathrm{ms}^{-1}$ and relative vorticity of $1 \cdot 10^{-5}\,\mathrm{s}^{-1}$ have been introduced. Since the algorithm assigns scores to its detections, an additional score threshold of 1150 was defined. These measures reduced the false alarm rate to 0.162. Allover, a detection scheme with a precision of 0.838 and a probability of detection of 0.455 was designed. Beyond that, the MDI algorithm proved to be suitable for estimating the strength of TC activity: the sum of the scores of individual hurricane seasons correlates statistically significantly ($r = 0.9$) with the accumulated cyclone energy (ACE). In general, the algorithm is recommended for detecting less specific anomalies that are still more unexplored in terms of their nature than TCs.

Zusammenfassung

Das Ziel dieser Mastarbeit ist die Detektion tropischer Zyklone (TCs) mithilfe eines multivariat anwendbaren, unüberwachten maschinellen Lernverfahrens – dem Maximally Divergent Interval-Algorithmus (MDI). Im Vergleich zu traditionellen Methoden der TC-Detektion, basiert dieses Verfahren nicht auf strikten, auflösungsabhängigen Grenzwerten von Variablen. Unter der Verwendung der ERA-Interim-Reanalyse und der TC-Datenbank IBTrACS wurde ein vorklassifizierter Datensatz entworfen, der die Verifikation der Algorithmus-Detektionen erlaubt. Der MDI-Algorithmus wurde auf die Region des Golfes von Mexiko während der Hurrikan-Saisons 2000 bis 2010 angewandt. Anfangs wurden die optimalen initialen Einstellungen des MDI-Algorithmus erarbeitet: Dabei zeigte sich, dass sich die Kullback-Leibler-Divergenz am besten eignet, um anormale Intervalle aufzuspüren. Des Weiteren sollte kein räumliches und zeitliches "Embedding" angewandt und die Detektionen sollten einen Zeitschritt nach vorne versetzt werden. Bei der univariaten Anwendung auf die relative Vortizität auf 850 hPa lieferte der Algorithmus mit einer mean Average Precision (mAP) von 0.537 die besten Ergebnisse. Eine multivariate Anwendung unter Einbezug anderer Variablen verbesserte die mAP nicht. Um die Fehlalarm-Rate zu minimieren, wurden weiche Grenzwerte der Windgeschwindigkeit ($8\,\mathrm{ms}^{-1}$) und relativen Vortizität ($1 \cdot 10^{-5}\,\mathrm{s}^{-1}$) eingeführt. Da der Algorithmus seine Detektionen entsprechend ihrer Anomalität mit einer "Score" versieht, wurde auch hier ein Grenzwert von 1150 definiert. Insgesamt konnte ein Detektionsalgorithmus mit einer Präzision von 0.838 und einer Detektionswahrscheinlichkeit von 0.455 entworfen werden. Darüber hinaus ist der MDI-Algorithmus fähig, die Stärke der TC-Aktivität abzuschätzen: Die Summe der Detektionsscores einzelner Hurrikansaisons korreliert statistisch signifikant ($r = 0.9$) mit der akkumulierten zyklonalen Energy (ACE). Insgesamt wird der Algorithmus zur Detektion weniger spezifischen Anomalien empfohlen, die noch unerforscht oder weniger gut definiert sind als TCs.

Contents

1 | Introduction

Natural hazards are a threat to humans and entire ecosystems. They can be divided into biological, geophysical, climatological, hydrological and meteorological events. 58 % of the worldwide financial loss caused by natural disasters between the exemplary period 2000 and 2007 can be attributed to windstorms. In absolute numbers this translates into more than US$47.5 billion annually. King et al. (2010) further estimate around 3500 annual fatalities from storms for this period, whereby all these numbers only include the direct effects of the storms. Indirect consequences, which are often caused by extreme precipitation, such as floods and landslides, as well as late effects such as epidemics are assigned to other categories, and thus the actual fatalities and damage amounts are substantially higher. In the meteorological sector, tropical cyclones (TCs) stand out as drastic natural disasters that are able to cause complex social, ecological and economic crises. The aforementioned number of fatalities is far exceeded by the estimates of Adler (2005), who assumes 10,000 people per year who died from the direct and indirect consequences of TCs in the period between 1993 and 2002. It is not only the sum of the storms, but also individual events that cause considerable damage: Pielke et al. (2008) estimate that hurricane Katrina caused damage amounting to US$80 billion in 2005. Munich RE (2019) even puts this figure at US$125 billion, which, after factoring out inflation, gives the event the same impact as the earthquake linked to the Fukushima nuclear disaster in Japan in 2011. Many other TCs, such as hurricane Harvey in the exceptionally active North Atlantic hurricane season 2017 (see Figure 1.1), are among

FIGURE 1.1: Picture of the three hurricanes Katia, Irma, and Jose (from left to right) at different states of their life cycle lying in a line over the Atlantic basin on September 6, 2017. On this day, the storm Katia over the Gulf of Mexico reached hurricane status. The eye of category 5 hurricane Irma was located north of Puerto Rico. Jose over the central Atlantic was also classified as a hurricane that day. The image is a mosaic of shots from several orbits and was taken with the Visible Infrared Imaging Radiometer Suite (VIIRS) mounted on the Suomi NPP satellite.

the most costly and deadly natural catastrophes of recent decades. Understanding, detecting, and predicting TCs has been and remains one of the most important tasks in averting fatalities and economic damage.

The extent to which TC activity is increasing with climate change is currently a controversial topic in research. The 5$^{\text{th}}$ Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) (2013) concluded that there is only *low confidence* that the activity of intensive TCs has increased since 1950. Merely in the North Atlantic it is *virtually certain* that the frequency and intensity of TCs have increased since 1970. However, the exact reason for this is still debated. There is also *low confidence* that humans are contributing to the observed changes. The reason for this is the lack of observational data and a lack of physical understanding of how anthropogenic drivers are linked to TC activity. Inconsistent study results also contribute to this ambiguous situation. Similar uncertainty applies to future changes in the first half of the 21$^{\text{st}}$ century. The IPCC (2013) also concluded that it is *more likely than not* that TC activity will increase locally in the second half of the 21$^{\text{st}}$ century, specifically in the Northwest Pacific and North Atlantic, depending on the climate change scenario. In addition, there are more frequent and extreme maritime floods, which have *likely* increased in the past since 1970 and will *likely* continue to do so in the future (IPCC 2013). As a result, for coastal regions that are already affected by TCs, the situation is expected to become more challenging (Walsh et al. 2016).

These findings are obtained through observations, reanalysis products and projections from climate models. The amount of data available at present is vast and cannot be evaluated manually by humans. To cope with today's data volumes, automated evaluation methods are needed. In the case of TCs, these are often traditional threshold methods that are triggered when a weather event shows certain wind speeds, vorticity values or temperature deviations (Walsh et al. 2007). There are many approaches with different threshold values based on quantities defined by experts (Tang and Monteleoni 2015). This makes the studies difficult to compare and the detectors often only applicable to a certain model resolution. Moderately resolved models with grid spacings of around 100 km are able to produce TCs (Knutson et al. 2010). TCs in such models are, however, typically not intensive enough and simulated with a too large spatial extent (Walsh et al. 2007). This thesis therefore explores whether the detection and understanding of TCs can be improved with the aid of machine learning.
With machine learning approaches, new methods are available to classify certain objects independently of previously determined characteristics. This is, for example, done by convolutional neural networks (CNNs), which are first applied to training data sets. They learn rules, mathematically expressed complex functions, which map input variables to different output variables or classes. For example, a ring-shaped pattern of high wind speeds enclosing a warm core indicates a TC by a high probability. Liu et al. (2016) have demonstrated the applicability of CNNs for TC detection: 98.9 % of TCs

were correctly identified by the algorithm. If these programs are trained using a data set in which both the instances to be classified and the remaining parts are available and labeled, one refers to this as *supervised machine learning*.

But even without these supervised procedures, machine learning techniques that identify non-specific anomalies should be able to detect TCs, as TCs leave footprints in variables such as wind speed, vorticity, temperature and humidity. An algorithm that is applied to reasonable variables, detects anomalies and ranks them according to their deviation from the rest of the data, should place TCs in the upper ranks. With this strategy, the algorithm does not need to be trained on a labeled data set beforehand, but can be applied *unsupervised* on data. As far as known, a *multivariate unsupervised machine learning algorithm* has not yet been applied for the detection of TCs so far, yet provides a promising new avenue. In this way, not only a detector could be developed, but also new variables or combinations of variables could be found, which would lead to an improvement in TC detection. This knowledge could in turn be used for traditional detection methods.
However, the requirements for an algorithm based on which the detector could be developed are high: since TCs extend over space and time, the algorithm must be able to detect anomalous intervals in spatio-temporal data. Many methods focus only on the detection of pointwise anomalies, for example, to identify individual defective products in a manufacturing process. In addition, the algorithm must be able to identify intervals of flexible length in time and space to consider the varying character of TCs. Furthermore, it should be applicable multivariately, i.e. it should not only be possible to analyze one variable such as wind speed, but several variables at the same time, since TCs cause fluctuations in different physical quantities. Finally, the algorithm has to be able to deal with large amounts of data time-efficiently, since calculation time on computing systems is expensive. All these requirements are met by the multivariate Maximally Divergent Intervals (MDI) algorithm designed by Barz et al. (2017). In this thesis, it will be used as detector of TCs in meteorological reanalysis data. As shown in Figure 1.2, the MDI algorithm was already tested in the study by Barz et al. (2017) in detecting hurricanes in a multivariate time series. For this purpose, observational data from a buoy near the Bahamas in 2012 were examined. Here, besides hurricanes, other anomalies were also detected. The aim now is to perform this detection process precisely and in spatio-temporal data sets.

Therefore, the central scientific question is: "Can a functioning detector for TCs be developed using the unsupervised MDI algorithm and what are the advantages and disadvantages compared to traditional methods?"

In *Fundamentals and theory* (Chapter 2), knowledge is gained about the classification, conditions, origin, characteristics and decay of TCs in order to get an idea of which variables the MDI algorithm should be applied to. The term *anomaly* is outlined and previous work that deals with the detection of TCs is highlighted.

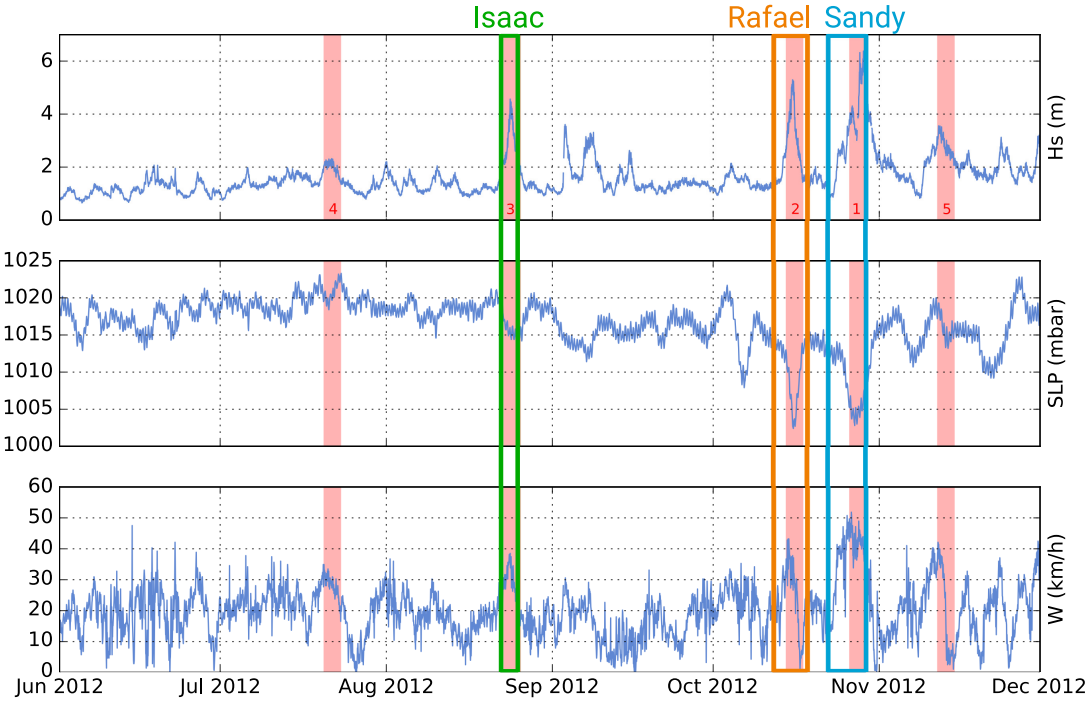


FIGURE 1.2: A multivariate time series of buoy measurement data near the Bahamas on which the MDI algorithm was applied for hurricane detection. The top row shows wave height, the middle row shows the pressure at sea level (PSL) and the bottom row shows wind speed. The red intervals represent the five highest ranked detections, with the respective rank indicated by small red numbers. The colored frames show the ground truths, i.e. the passage of hurricanes near the measurement site. Reprinted from Barz et al. (2017).

In order to understand how the MDI algorithm performs in detecting TCs, it is applied to reanalysis data from the past, which are realistic representations of the former atmospheric state on a regular grid. Since meteorological events are not labeled in these data, a data set being as complete as possible, containing the positions and times at which past TCs occurred, is needed. These data sets are presented in *Data and methods* (Chapter 3). In addition, the way the MDI algorithm works and metrics that can be used to evaluate the detection skill are explained.

Based on these data sets, the detections provided by the MDI algorithm are evaluated in *Results* (Chapter 4). It is attempted to improve its detection skill step by step by preprocessing the data, changing initial settings of the algorithm and post-processing of the detections.

In the *Discussion* (Chapter 5), the final steps are to assess the detection skill and look for reasons for success or failure of the method. Final recommendations are made in *Conclusions and Outlook* (Chapter 6).

2 | Fundamentals and theory

This section covers the fundamentals of tropical cyclones (TCs) in Section 2.1 and the definition of the term *anomaly* as well as the presentation of existing anomaly detection methods in Section 2.2. In Section 2.3, the focus lies on the state of the art of TC detection.

2.1 Tropical cyclones

The detection of tropical cyclones (TCs) requires basic knowledge of their structure, dynamics and formation. This knowledge allows to define the variables to which the algorithm should be applied and what the physical boundary conditions are in order to improve the detection. The following section therefore provides and overview of the classification (Section 2.1.1) and basic physical concepts for the life cycle of a TC: this includes formation (Section 2.1.2), intensification (Section 2.1.3), full-grown status

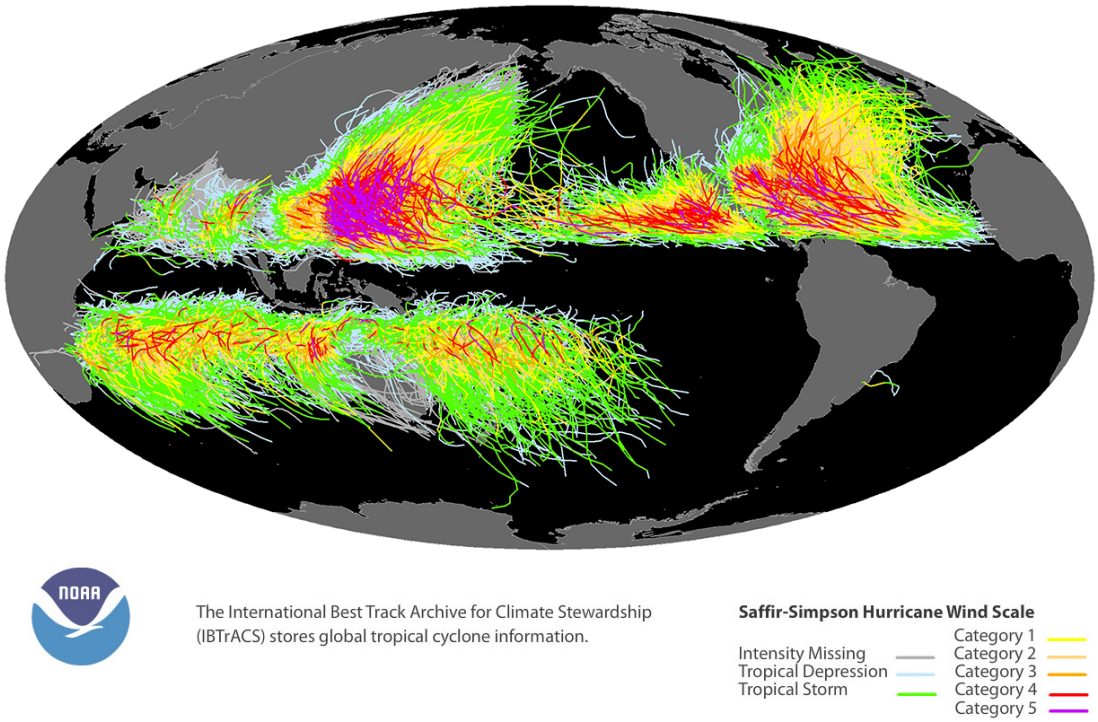


FIGURE 2.1: Global map of all TCs in the International Best Track Archive for Climate Stewardship (IBTrACS). For further explanations regarding this database, see Section 3.2. The cyclones are colored according to their intensity on the Saffir-Simpson scale. In this map, TCs in the time period from 1850 to 2017 are shown. Retrieved from National Centers for Environmental Information (NCEI) (2017).

(Sections 2.1.4 to 2.1.6) up to its decay (Section 2.1.7).

2.1.1 Classification

TCs are atmospheric vortices that form over warm tropical oceans. Each year, about 80 to 90 TCs occur (e.g. Frank and Young 2007; Gray 1985), 80 % to 90 % of them within 20° latitude north and south of the equator. As can be seen in Figure 2.1, TCs occur over all tropical oceans, but there are remarkable local accumulations and tropical regions with little activity. Depending on the region, these weather phenomena are termed and classified differently. However, the classification is commonly based on their *intensity*, i.e. the 1- or 10-minute maximum sustained wind speed. In this thesis, the classification of the National Hurricane Center (NHC) is chosen (see Table 2.1) which is based on the Saffir-Simpson scale (Simpson and Saffir 1974).
At the beginning of the development there is a tropical depression, which according to definition has a wind speed of less than $63\,\mathrm{kmh}^{-1}$ sustained for 1 min. It can evolve into a tropical storm (TS) of up to $118\,\mathrm{kmh}^{-1}$. Above this, the storm is classified as a hurricane, and from $178\,\mathrm{kmh}^{-1}$ upwards as a major hurricane, which falls into category 3 on the Saffir-Simpson scale. In the Indian and southern Pacific Oceans, the term *hurricane* is not common, in the northwestern Pacific it is referred to as *typhoon*.
To quantify the strength of a TC or an entire hurricane season, the accumulated cyclone energy (ACE) (Bell et al. 2000) is commonly used. It is the sum of all squares of the maximum sustained wind speed in knots at the times t 0 UTC, 6 UTC, 12 UTC and 18 UTC. To avoid large numbers, it is scaled with 10^{-4}:

$$\mathrm{ACE} = 10^{-4} \sum_{t} v_{max}^2(t) \tag{2.1}$$

For a single storm, the sum is calculated over its lifetime, for a whole season over all storms of the investigated area which could be one ocean basin for instance.

2.1.2 Large-scale conditions for TCs

In this section it is addressed which large-scale external conditions are necessary for TC formation. This knowledge is advantageous for the application of the algorithm, since the detection is not carried out by hard limits in this study, for example in wind speed, but by the presence of anomalies. Anomalies in variables which individually do not necessarily indicate a TC, but in combination with anomalies in other variables more clearly indicate the presence of a TC, can thus be considered. The latter idea is implemented in the algorithm and used by applying it to multivariate data sets.
The formation of a TC involves many processes that take place on convective to synoptic scales (Tory and Frank 2010). There are some basic conditions needed for TC formation that Briegel and Frank (1997) summarized based on previous studies (e.g. Gray 1967, 1979, 1985):

(i) Sea surface temperatures above 26.5 °C to 27 °C in combination with an ocean mixed layer depth of about 50 m

TABLE 2.1: The classification of TCs according to Simpson and Saffir (1974) and National Hurricane Center (NHC) (2012).

	1-minute sustained winds		
Beaufort scale	in kn	in km h^{-1}	Classification
0–7	<34	<63	Tropical Depression
8	34–63	63–118	tropical storm (TS)
12+	64–82	119–153	Category 1 hurricane
	83–95	154-177	Category 2 hurricane
	96–112	178–208	Category 3 major hurricane
	113–136	209–251	Category 4 major hurricane
	≥ 137	≥ 252	Category 5 major hurricane

(ii) A deep, conditionally unstable atmospheric layer

(iii) High cyclonic absolute vorticity in lower tropospheric levels

(iv) Organized deep convection in a region where on average air rises and where there is high humidity in the middle troposphere

(v) Weak to medium vertical wind shear

These conditions are not completely independent of each other and are present over large parts of the tropical oceans in summer. Since TCs are only occasionally occurring phenomena, further triggers for TC formation are necessary. Moreover, for points (iii)-(v) it is not clear to what extent these are initial conditions or consequences of TC formation.

Analyses by McBride and Zehr (1981) and observational case studies such as by Zehr (1992) have shown that the storms mainly form where at the same time anticyclonic relative vorticity occurs in upper levels and cyclonic relative vorticity in low levels. A weak eastern wind shear between 850 and 200 hPa is also favorable for storm formation (Kurihara and Tuleya 1981). The theory that high humidity is required in the low and medium troposphere is supported by Bister and Emanuel (1997). This humidity prevents the blocking effect of downdrafts on the forming circulation (see Section 2.1.3).

Especially monsoon troughs seem to positively influence TC formation. Gray (1967) states that a large portion of TCs form near these troughs. These are folds of the intertropical convergence zone (ITCZ), for example, towards the north, where southwesterly winds blow from the south and northeasterly winds from the north, which then converge. This configuration produces low-level vorticity, which is beneficial to the formation of TCs.

Another amplifying factor is the presence of tropical waves. In monsoon troughs, TCs can also form without the aid of tropical waves, but Roundy and Frank (2004) suggest that the interaction of monsoon troughs and tropical waves is the most common mechanism for TC formation. Tropical waves propagate zonally along the equator. Observations suggest that equatorial and near-equatorial waves favor formation of TCs by

enhancing deep convection and vertical mass flow, increasing low-level vorticity, and by modifying circulations in a supportive manner.
The connection between tropical waves and formation of TCs was first discovered in observations of African waves. African waves are perturbations of low air pressure that originate above the Sahara and propagate westward. They produce about half of the hurricanes in the northern Atlantic (Frank and Clark 1980).
Tropical waves can also create an environment that protects the growing TC from external factors, for example by creating a flow of closed streamlines relative to the zonal flow. As a metaphor, these waves are referred to as *marsupial pouch*, which protects the *TC embryo* from unfavorable factors such as dry air or wind shear. As soon as a core vortex has developed that is strong enough to withstand external factors, the storm separates from its mother wave and propagates independently (Dunkerton et al. 2008). Polewards of 20° other factors become dominant for TC formation. The storm genesis in the extratropical Atlantic is often caused by troughs in the upper atmosphere and subtropical decaying frontal zones (McTaggart-Cowan et al. 2008). As in tropical waves, the increased lower tropospheric vorticity and the induced upward movement on synoptic scales favor TC formation. According to Davis and Bosart (2003), strong mid-latitude cyclones can also undergo this transition, transforming from a cold core vortex to a warm core vortex.

2.1.3 Intensification

At the stage of a tropical depression (see Section 2.1.1) the storm can sustain itself, because the system obtains more energy from surface sensible heat and latent heat from condensation of moisture than it loses energy through friction (Rotunno and Emanuel 1987). Mesoscale convective systems (MCSs) play a crucial role in this. As shown in Figure 2.2 (a)-(b), MCSs consist of a narrow area of convective precipitation and a larger area of stratiform precipitation. Such MCSs are self-sustaining and can remain active for several days. In the convective region, convergence occurs in the lower to middle troposphere and divergence above it. In the zone of stratiform rain, convergence prevails in the middle troposphere and divergence below and above. The convergence of the convective zone promotes the strengthening of vorticity but competes with the divergence of convective downdrafts and the divergence below the stratiform zone. This negative feedback, in which convective systems prevent themselves from intensifying through downdrafts and subsidizing divergence, is often referred to as *kinematic roadblock*.
A requisite for sufficient intensification is convective warming by latent heat release, which takes place in the convective area of a MCS. Rising air masses initially cool down due to the decrease in pressure, but at the same time water vapor condenses once saturation is reached and latent heat is released. This reduces the cooling, so that the ascending air parcel is warmer than the ambient air and can rise further until it diverges in the upper troposphere. This shifts the divergent area to the upper atmosphere, causing an inflow that extends across the entire boundary layer. This in-up-out path of the

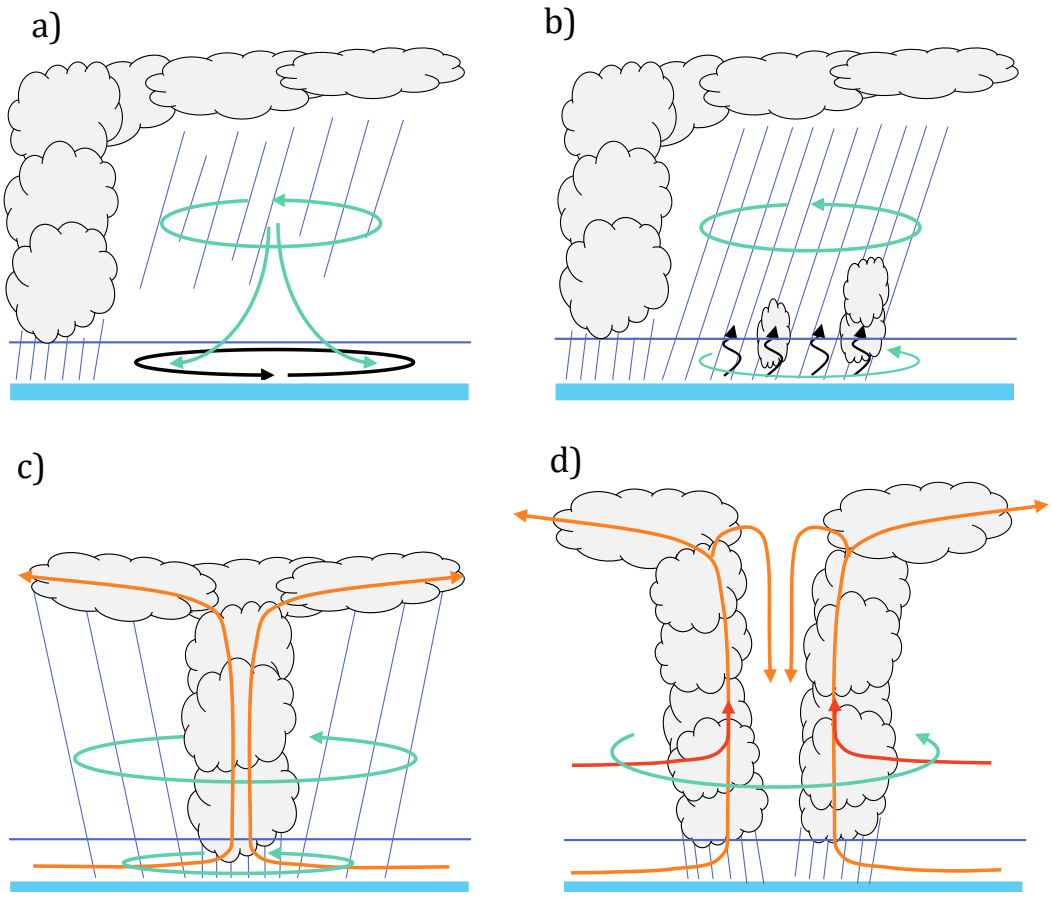


FIGURE 2.2: Formation of a TC from a MCS and final circulation pattern. At the bottom the ocean surface is indicated in blue, the boundary layer is located above with the boundary layer top depicted by a blue line. a) The stratiform precipitation of a MCS cools and moistens the upper part of the lower troposphere. As a dynamic response, the mid-level cyclonic circulation (turquoise) intensifies. Subsiding air dries and warms the lower atmosphere. b) The mesocyclone (turquoise arrows) propagates into the boundary layer. The enhanced wind speeds increase the interaction between sea and air and result in a moistening of the boundary layer. c) With the cooler upper part of the lower troposphere convection can build up. In addition, the partially stable stratified middle troposphere prevents downdrafts from blocking the boundary layer convergence. d) In the interior there is subsidence of air, which creates a cloudless, warm eye. In the boundary layer inflow (lower orange inflow branch), the angular momentum M is not conserved due to friction, but it is conserved above it (red inflow branch). Reproduced after Bister and Emanuel (1997) and Smith and Montgomery (2016).

air later develops into the *secondary circulation*. Without convective heating, only a shallow secondary circulation would form, with an inflow of air in the boundary layer and a divergence directly above.

In reality, there are also MCSs that do not have one convective and one stratiform region, but several of them (Houze 2004). Moreover, several MCSs can be present at the same time together. A complex of several MCSs in different stages, initially unorganized, can begin to interact (e.g. Simpson et al. 1997). This is particularly the case when the convective complex is embedded in a protected environment such as tropical waves (Tory and Frank 2010).

In simulations by Hendricks et al. (2004) and Montgomery et al. (2006), these many small-scaled and often short-lived convective cells within the MCS(s) ultimately lead

to an intensification of the secondary in-up-out circulation. This circulation can be described simplified by the Eliassen balanced vortex model (Eliassen 1951). The circulation develops in an idealized axis symmetric way and in gradient wind as well as hydrostatic balance. In Montgomery et al. (2006) the circulation equation is given as follows:

$$\frac{\partial}{\partial r}\left(\frac{N^2}{r}\frac{\partial \psi}{\partial r} - \frac{\xi}{r}\frac{\partial v}{\partial z}\frac{\partial \psi}{\partial z}\right) + \frac{\partial}{\partial z}\left(-\frac{\xi}{r}\frac{\partial v}{\partial z}\frac{\partial \psi}{\partial r} + \frac{I^2}{r}\frac{\partial \psi}{\partial z}\right) = \frac{\partial Q}{\partial r} - \frac{\partial}{\partial z}(\xi F) \tag{2.2}$$

Here ψ is the stream function of the toroidal circulation with $u = -\frac{1}{r}\frac{\partial \psi}{\partial z}$ and $w = \frac{1}{r}\frac{\partial \psi}{\partial r}$. v is the tangential, axisymmetric wind, which is also termed *primary circulation*. Furthermore, $\xi = (f + 2v/r)$ with the Coriolis parameter f, and $I^2 = \xi\eta$ denotes the inertial stability parameter, which, similar to the Brunt-Väisälä frequency N^2, describes the stability of an air parcel in the given circulation respectively stratification. z and r are the coordinates in the cylindrical system, Q describes the diabatic heating and F the momentum forcing terms. Under the assumption that $\frac{\partial v}{\partial z}$ and F are negligible and N^2 and I^2 are positive and almost constant (radially and vertically), the result is:

$$N^2\frac{\partial w}{\partial r} - I^2\frac{\partial u}{\partial z} \approx \frac{\partial Q}{\partial r} \tag{2.3}$$

In the case of the maximum diabatic heating rate in the center of circulation, the result is $\frac{\partial Q}{\partial r} < 0$, a negative $\frac{\partial w}{\partial r}$ and a positive $\frac{\partial u}{\partial z}$. Altogether, the secondary circulation consists of an inflow at the surface, an ascent near the heating source (see Figure 2.3) and subsidence on the outside.

The consequences for the primary circulation, i.e. the tangential axisymmetric wind, can be described by the tangential wind tendency equation. This equation is valid in a TC according to Kepert (2010). Assuming axial symmetry, it can be expressed in cylindrical coordinates as follows:

$$\frac{\partial v}{\partial t} = -\left(u\eta + \omega\frac{\partial v}{\partial p} - F_\lambda\right) \tag{2.4}$$

In this equation, η is the vertical component of the absolute vorticity and F_λ is the frictional force in the direction of the tangential wind. p is the pressure.

An inflow $-u$ increases the tangential wind v. The primary circulation shown in Figure 2.3 is thus also built up as a result of the increasing secondary circulation. Furthermore, for axisymmetric circulations where the friction is small and the radial winds are weak, the gradient wind balance applies:

$$\frac{1}{\rho}\frac{\partial p}{\partial r} \approx fv + \frac{v^2}{r} \tag{2.5}$$

This is the equilibrium of Coriolis force, centrifugal force and radial pressure gradient force. In a system that develops in this equilibrium, the radial pressure gradient $\partial p/\partial r$ must also change whenever the tangential wind speed v changes. This is achieved by diabatic heating inside the storm, but also by descending, warming air masses outside the heating region. As a consequence of the warm core in the middle troposphere and

the gradient wind balance, the wind speed above the boundary layer decreases with height.
All in all, in this model explains the increasing primary circulation, which reaches its maximum wind speed in the vicinity of the heat source.

As already mentioned, MCSs under non-optimal conditions still prevent themselves from intensifying, since the divergence of downdrafts crucially weakens the low-level vorticity. Bister and Emanuel (1997) propose the following concept of how a TC can form from an existing MCS and overcome the kinematic roadblock, as sketched in Figure 2.2:

(a) The precipitation of the stratiform zone cools and moistens the upper part of the lower troposphere. As a result, this part of the troposphere becomes stably stratified and saturated. The cooling is dynamically compensated by an intensification of the vortex in the middle troposphere.

(b) This cyclonic vortex of the middle troposphere propagates into the lower atmosphere.

(c) The higher winds speeds lead to increased heat and moisture fluxes between the ocean and the lower atmosphere. In addition, the cool upper part of the lower atmosphere allows convection to build up originating from the still unstable bottom layer. This convection has a lower downdraft potential as the atmosphere is partly stably stratified as described in (a).

An ideal environment for the formation of TC should therefore include a shallow, unstable bottom layer and a moisture-neutral, saturated layer above (Tory and Frank 2010). According to Bister and Emanuel (1997), a MCS can create this environment and thus form a storm center. This theory is also referred to as *top-down showerhead*.

2.1.4 Concept of conservation of angular momentum

An important concept for the further spin-up inside a radial storm is the conservation of angular momentum. Assuming an axisymmetric storm and the negligence of friction, the tangential momentum equation reduces to the statement that the angular momentum per unit mass, M, is conserved when rings of air parcels move around on a meridional plane (Smith and Montgomery 2016). To illustrate this, the simplified version of the tangential momentum equation

$$M = rv + \frac{1}{2}fr^2 \tag{2.6}$$

is transformed into

$$v = \frac{M}{r} - \frac{fr}{2}. \tag{2.7}$$

If a ring of air parcels reduces its distance to the storm center and M remains constant, the tangential velocity v increases.

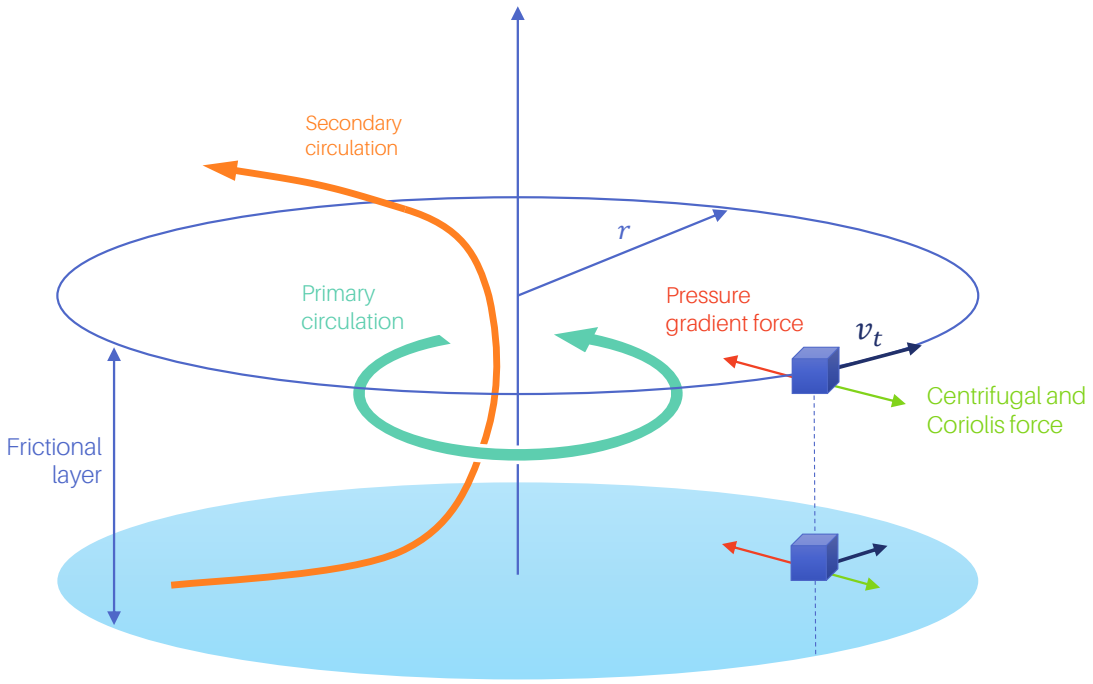


FIGURE 2.3: Illustration of the primary and secondary circulation pattern within a TC. In addition, the forces acting on an air parcel in and above the frictional layer are shown. In the case of the lower air parcel, the friction reduces the centrifugal and Coriolis force. This results in a net inward force. Reproduced after Smith and Montgomery (2016).

2.1.5 Structure and dynamics of a mature TC

Mature storms have an almost axisymmetric circulation in their central area. The strongest winds occur at the edge of the eye, where comparatively calm conditions are prevailing. Typically, the eye of a TC has a diameter of 20 to 100 km (Kepert 2010). Compared to the ambient environment, the pressure inside the eye can drop by up to 10 %. From a hydrostatic point of view, this requires the storm to have a warm core that extends vertically over almost the entire troposphere (Haurwitz 1935). The *eyewall*, a ring of deeply convective clouds, encloses the eye completely or partially and extends from the sea surface to the tropopause, widening with increasing height.
In the frictional boundary layer, which is about 1 km deep, and in the outflow layer, the gradient wind equation 2.5 is not valid. Therefore, the primary axisymmetric circulation is superimposed by the secondary circulation, which consists of inflow, upflow in the eyewall and the spiral *rainbands* and outflow. The latter comprises a thin layer below the tropopause. As illustrated in Figure 2.3, the secondary circulation inflow also results from the fact that friction offsets the gradient wind balance, allowing air masses to flow to the storm center. The inflow interacts with the sea by momentum transfer, absorbs moisture and provides it to the convective center (Ooyama 1969). The release of latent heat in the eyewall causes the air masses to subside in the storm center, leading to the formation of the warm core and the cloudless eye of the storm. Deep convection, which occurs in the eyewall and extends throughout the full vertical extent of the troposphere, is important here: without it, an outflow would take place directly above the boundary

layer, whereby the air parcels would flow to the outside while maintaining their angular momentum M. Ultimately, M would dissipate in this system, since it is not conserved in the inflow of the boundary layer. As a result, the storm's rotation would slow down. In other words, convection must be strong enough to ventilate the converging flows of the boundary layer in a way that they do not diverge directly above the boundary layer, but allow an inflow in the middle troposphere with preserved M (Ooyama 1969), as shown in Figure 2.2 (d).

According to this classical theory, the tangential wind speed in the middle troposphere is greater for a given radius from the center than in the boundary layer below. However, this is not supported by simulations and observations (Smith and Montgomery 2016). Instead, the highest tangential wind speed occurs at the boundary layer top in the inner core (Smith and Vogl 2008). With Equation 2.7 it is possible to heuristically understand this process: as a result of friction, M is not conserved in the boundary layer and is reduced with time. Smith and Montgomery (2016) describe the following mechanism: "[n]evertheless, if, for an air parcel spiraling inwards in the boundary layer, the relative rate at which M is reduced by friction is less than the relative rate at which the parcel's radius decreases, then v will increase". This is the case with a strong inflow as a result of deep convection. At a certain point, the tangential wind speed becomes very high and the resulting centrifugal force produces a net force directed outwards. At this point the inflow slows down and the air rises to form the eyewall where the highest wind speeds take place. They usually occur at a height of 300 to 800 m (e.g. Franklin et al. 2003; Kepert 2006a,b). Below this *low-level jet*, the wind speed in the lower 100 to 200 m increases logarithmically with height (Franklin et al. 2003; Powell et al. 2003).

In the upper part of the storm, where the outflow occurs below the tropopause, the circulation is less stable, which can lead to pronounced asymmetries, as one can see in Figure 1.1, for example, from hurricane Katia over the Gulf of Mexico. In general, stronger storms tend to be more symmetric.

In mature storms, eyewall replacements can occur, in which a second eyewall forms at a larger radius. The second eyewall counteracts the already existing eyewall by causing air to sink over the primary eyewall (Willoughby et al. 1982) and cutting off the supply of air with high enthalpy from the inflow to the inner eyewall. Often the inner eyewall dissipates and is replaced by the outer eyewall, which contracts and intensifies. This process is called the *eyewall replacement cycle* and leads to a fluctuating intensity of the storm.

Another typical feature of TCs are outward spiraling rain and cloud bands, which were discovered early in satellite imagery (e.g. Maynard 1945; Wexler 1947) and are also visible in Figure 1.1. They contain both stratiform and convective clouds. By consuming enthalpy rich inflow, they can weaken the storm overall. Moreover, the latent heat released in them does not intensify the storm, as the heat release outside the storm core is not efficient. On the other hand, the bands protect the storm interior from environmental wind shear and act as a buffer (Kepert 2010).

So far, the storm was examined in an idealized, undisturbed environment. This is not

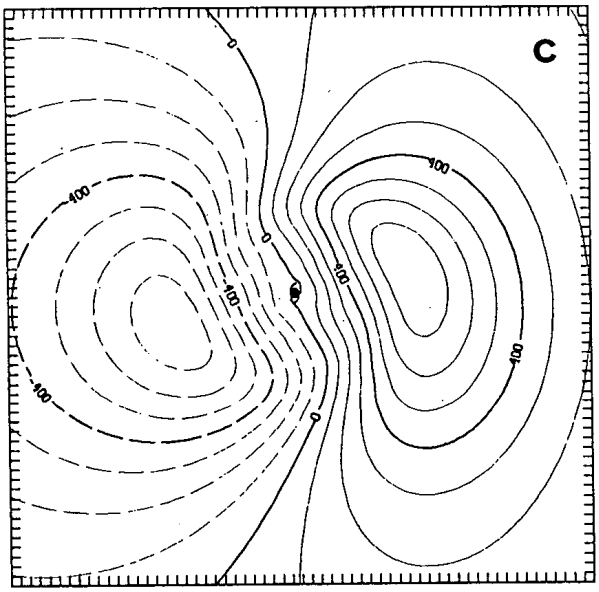


FIGURE 2.4: Result of a 24 h integration of Equation 2.9. The stream function of the asymmetric field of relative vorticity is shown with a contour spacing of $1 \cdot 10^5\,\mathrm{m^2 s^{-1}}$, the unit of annotated numbers is $1 \cdot 10^4\,\mathrm{m^2 s^{-1}}$. Solid lines show anticyclonic, dashed lines show cyclonic stream functions. Reprinted from Fiorino and Elsberry (1989).

the case in reality, as external factors of atmosphere, ocean and land have an influence on the storm. In the case of atmospheric wind shear, it has been shown that there is a threshold above which it weakens a mature storm. This is because of shifting the upper from the lower circulation and finally decoupling it. According to Kepert (2010), the wind shear threshold above which the storm is weakened is 8 to $10\,\mathrm{ms^{-1}}$ between the upper and lower troposphere in simulations. This is also confirmed by observations from DeMaria and Kaplan (1994, 1999), DeMaria et al. (2005), and Gallina (2002). Shear also leads to asymmetries in the distribution of clouds, precipitation and convection in a TC (Frank and Ritchie 1999, 2001; Wang and Holland 1996; Wong and Chan 2004).

2.1.6 Movement

The lifespan of TCs generally ranges from one or two days up to several weeks (Kepert 2010). During this time, the storm often moves characteristically as visible in Figure 2.1. In a first approximation, the vortex is advected by the surrounding large-scale flow. As in a river, the vortex is moved and steered downstream. Mathematically this can be expressed as follows:

$$\frac{\partial \zeta}{\partial t} = -\mathbf{V} \cdot \nabla \zeta \tag{2.8}$$

ζ is the relative vorticity, $\mathbf{V}$ the ambient flow and ∇ the two-dimensional, horizontal nabla operator. This concept has proven to be particularly useful in short-term predictions. A practical problem, however, is to determine $\mathbf{V}$ precisely. Particularly in non-barotropic regions, the question arises over which atmospheric layers the winds should be averaged in order to estimate the ambient flow (Chan 2010). A main result of the studies dealing with the estimation of $\mathbf{V}$ is that only the ambient flow by itself cannot predict the TC track (Chan 2010). There is a difference of 1 to $2.5\,\mathrm{m\,s^{-1}}$ between the TC movement and the steering flow. In addition, TCs move deviating further to western and polar directions compared to $\mathbf{V}$. Equation 2.8 must therefore be extended

by another term:

$$\frac{\partial \zeta}{\partial t} = -\mathbf{V} \cdot \nabla \zeta - \beta v_m \tag{2.9}$$

Here β is the latitude-dependent variability of the Coriolis parameter and v_m the meridional wind. This *beta effect* acts as follows: due to its cyclonic rotation, the TC advects relatively high planetary vorticity from poleward regions to its western side and generates a cyclonic circulation. In addition, relatively low planetary vorticity is advected to the eastern side of the TC resulting in an anti-cyclonic circulation. The resulting asymmetric vorticity field and its circulation streamlines are shown in Figure 2.4. The resulting vorticity field causes a shift of the TC towards the west and towards the pole. The movement of a TC is therefore a combination of the steering flow and the beta effect. This is true in barotropic atmospheric conditions.
In Figure 2.1 one can recognize how the tracks of the TCs initially start in the tropics and then curve towards the poles and west, mainly due to the beta effect. Thereafter, the influence of the westerly wind zone dominates, deflecting the storm trajectory to the east.

2.1.7 Landfall, decay and extratropical transition

When a storm moves from the ocean over islands or continental landmasses, this is referred to as *landfall*. This is where typically the greatest impact occurs on humans, who otherwise only experience the direct influence of TCs on oil platforms, ships or in air traffic. The storm is weakened by two mechanisms: firstly, the supply of latent and sensible heat through the sea surface is cut off. This weakens the convection in the storm core, resulting in a slowdown of the secondary circulation and less subsidence of air masses in the eye of the storm. This cools the warm core as a whole and causes the air pressure in the storm center to rise. Consequently, the wind speed of the primary circulation also decreases (see Equation 2.5). Secondly, the higher friction of the rough land surface weakens the winds in the boundary layer (Laing and Evans 2011).
The characteristics of the coastal areas and the interior of the landmass are crucial. Relatively shallow, humid areas such as swamps continue to provide latent heat fluxes to the storm to some extent. Built-up or forested areas weaken the storm by their roughness and the reduction in moisture supply. In these areas, however, strong turbulence and channelling effects can cause gusts that temporarily exceed the storm's large-scale wind speed and cause greater damage. Downdrafts are also more frequently mixed downwards by a turbulent boundary layer (Franklin et al. 2003).
If the storm moves back over the tropical, open ocean after its landfall, re-intensification may occur. Another pathway a TC can take is the extratropic transition. TCs moving towards the pole transform often into high-impact mid-latitude weather systems with heavy rainfall, high wind speeds and large wave heights (Harr 2010). At the transition, the TC usually loses its symmetry and its intensity decreases. Reasons for this are the lower sea surface temperature (SST) and the increased wind shear in mid-latitudes (e.g.

Frank and Ritchie 2001; Klein et al. 2000; Thorncroft and Jones 2000). However, this often increases the area of precipitation which can then affect larger areas (Evans and Hart 2008).

2.2 Anomalies and their detection

The term *anomaly* is not strictly defined and can be outlined, for instance, in the words of Chandola et al. (2009):

> "Anomalies are patterns in data that do not conform to a well-defined notion of normal behavior."

The task of anomaly detection is thus to detect these anomalous patterns in the "notion of normal behavior". If focusing on machine learning, there are three different approaches: *supervised*, *unsupervised* and *semi-supervised machine learning*. In supervised learning there is a data set with labeled nominal and anomalous fragments, also known as ground truths. The algorithm acquires a set of rules based on the labeled data and uses them to classify unknown data sets (Kotsiantis et al. 2007). In the case of unsupervised algorithms, this labeled data set is not available for training purposes and the data to which it is applied is entirely unknown to the algorithm. In semi-supervised learning, only the nominal data is available; the algorithm must therefore recognize the deviating instances in the actual application by itself (Chandola et al. 2009).
In the specific application of these algorithms to atmospheric variables, the definition of Hawkins (1980) is adequate, who describes an anomaly as "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism". Here, this "different mechanism" is the genesis of a TC that leads to weather extremes in temporal and spatially limited intervals that deviate strongly from the nominal annual cycle. The MDI algorithm, which searches for intervals in spatial and temporal data sets that deviate the most from the distribution of the variables in the remainder of the data set, adopts this idea. Furthermore, it does not search for deviating point anomalies, but for intervals whose size can even be flexible. In addition, the MDI algorithm can be applied not only to one variable, but to several variables simultaneously, such as temperature, pressure and humidity (for further details on the MDI algorithm, see Section 3.3).
There are hardly any approaches to anomaly detection that cover the full range of the MDI algorithm capabilities: Keogh et al. (2005) and Ren et al. (2018) follow a similar idea and compare data blocks with the rest of the time series. However, their method can only be applied to univariate data and the length of the intervals must be determined in advance. Liu et al. (2013) compare two intervals of fixed size using the Kullback-Leibler (KL) or Pearson divergence to find change-point anomalies at which a data series changes its distribution permanently. However, this method does not work when searching for intervals of variable size and the basic objective is different because intervals are not compared with all the remaining data. Wu et al. (2008) pursue an

approach in which anomalies are first found in space and then in time and subsequently merged. However, this method can only be applied to binary data and is costly in terms of computing time (Barz et al. 2018).

2.3 Existing TC detectors

There are already several detectors for TCs. Their approaches, strengths and weaknesses are presented in this section. The detection of TCs is basically done in two steps: first, the detection of single grid points or intervals that could represent a TC, referred to as *candidate detection*, and second, the connection of single spatial instances to a storm, the so-called *tracking*. This study concentrates mainly on the first step, which is why no time specifications are given below as to how long certain thresholds must be maintained. This is usually only important as criterion in the tracking step.
The presentation of the detectors is done chronologically because, as it is often the case with the development of scientific tools, previous methods have been improved, adapted and combined. A complete list of all TC detectors up to 2017 can also be found in Ullrich and Zarzycki (2016). In tabular form, Walsh et al. (2007) summarize the various studies published at that time.
One of the first approaches dates to Bengtsson et al. (1982), who identified TCs in a one-year $\sim 200\,\mathrm{km}$ resolved forecast model. Only events occurring at a geographical latitude of $< 30°$ are considered TCs. Furthermore, the 850 hPa wind speed must be $> 25\,\mathrm{ms}^{-1}$ and the 850 hPa relative vorticity must be greater than $7.5 \cdot 10^{-5}\,\mathrm{s}^{-1}$.
Broccoli and Manabe (1990) identified TCs in two models with $\sim 600\,\mathrm{km}$ and $\sim 300\,\mathrm{km}$ resolution respectively. The pressure at sea level (PSL) requires a local minimum of 1.5 hPa or 0.75 hPa in the coarser or finer resolution. The local wind speed needs to be $> 17\,\mathrm{ms}^{-1}$ and the latitude $< 30°$.
Wu and Lau (1992) detect TCs in a 7.5° longitude × 4.5° latitude model. A local minimum must be present at the 1000 hPa geopotential height. Furthermore, a positive 950 hPa relative vorticity, a negative 950 hPa divergence and a positive 500 hPa vertical velocity is required. The latitude limit is $< 40.5°$. The wind speed at 950 hPa must exceed $17.2\,\mathrm{ms}^{-1}$. It should be emphasized that for the first time a warm core criterion was applied: the layer thickness between 200 and 1000 hPa has to exceed the surrounding average layer thickness by 60 m.
Haarsma et al. (1993) also pursue this idea. They track TCs in a $\sim 300\,\mathrm{km}$ resolution data set. In addition to the already presented thresholds for PSL and 850 hPa relative vorticity, they provide several temperature anomaly criteria: $\Delta T_{250} > 0.5\mathrm{K}$ at 250 hPa, $\Delta T_{500} > -0.5\,\mathrm{K}$ at 500 hPa, and $\Delta T_{250} - \Delta T_{850} > -1.0\,\mathrm{K}$.
Bengtsson et al. (1995, 1996) use similar wind speed and relative vorticity thresholds at 850 hPa. They apply the detector to a T106 ($\sim 125\,\mathrm{km}$) data set. The warm core criterion is realized by a sum of the anomalies: $\Delta T_{700} + \Delta T_{500} + \Delta T_{300} > 3\,\mathrm{K}$ and $\Delta T_{300} > \Delta T_{850}$. This makes the algorithm more flexible in cases where a warm core is not exactly located near one of the expected levels like in Haarsma et al. (1993).
Tsutsui and Kasahara (1996) scan for TCs in a T42 ($\sim 300\,\mathrm{km}$) data set. A new cri-

terion is the consideration of 900 hPa relative vorticity and divergence, as well as the layer thickness between 200 hPa and 1000 hPa. Furthermore, they introduce a maximum topographic height of 400 m above which the TC can be located. This excludes inland storms with TC-like wind speeds. A minimum precipitation of 100 mmd^{-1} is also introduced as a threshold value.

The studies by Vitart et al. (2001, 1997, 1999) and Vitart et al. (2003) are based on the previous ideas regarding threshold values for variables. However, the specified thresholds are used as a basis in many other studies (e.g. Knutson et al. 2007; Zhao et al. 2009), which is why they are explicitly mentioned here: $vo_{850} > 3.5 \cdot 10^{-5}\,\mathrm{s}^{-1}$ in combination with a nearby PSL minimum. Between 500 and 200 hPa, a temperature anomaly of at least 0.5 K must be present, as well as a local maximum of at least 50 m in the layer thickness between 1000 and 200 hPa. The wind speed must be at least 17 ms^{-1} at one point. Camargo and Zebiak (2002) point out that the detection schemes of Vitart et al. (1997) and Bengtsson et al. (1995) do not spot some visually evident TCs.

Walsh (1997), Walsh and Katzfey (2000), and Walsh and Watterson (1997), working with a 125 km resolved regional climate model for Australia, impose the condition that the tangential wind speed at 850 hPa must be greater than that at 300 hPa in addition to the established thresholds. This is because the highest wind speeds of TCs occur in the boundary layer (see Section 2.1.5).

All previous approaches applied *absolute thresholds*, i.e. they are uniform even when applied to global models. This is problematic because TCs have different properties in each ocean basin, for example because they can develop undisturbed for longer periods in the Northwest Pacific, whereas TCs in the Atlantic make landfall earlier. Camargo and Zebiak (2002) therefore use the approach of Bengtsson et al. (1995, 1996), except that the thresholds in their T42 climate data set have been adapted individually for each ocean basin. This approach is referred to as geographical-adapted *relative thresholds*.

When reviewing the previous wind speed thresholds, it is noticeable that the threshold values vary considerably. The same applies to the relative vorticity, which is the curl of the horizontal wind speed and is therefore directly linked to it. These different thresholds are necessary because TCs show different intensities depending on the model resolution. With coarser resolutions, lower wind speeds are modelled. While the adjustment of thresholds is necessary on the one hand, it also reduces the comparability of the studies on the other hand. Horn et al. (2014) thus state that the different wind speed thresholds are one of the main reasons for the varying assessment of TC activity of the detectors. One milestone regarding this problem is the study of Walsh et al. (2007). In the study a threshold for wind speeds is elaborated which is approximately linearly dependent on the model resolution and increases with increasing model resolution. In many further studies on TC detection (e.g. Caron et al. 2013, 2011; Horn et al. 2014) this idea was applied. However, Murakami (2014) argues that the method of Walsh et al. (2007) underestimates the TC number in multiple reanalysis data sets.

Murakami and Sugi (2010) advance this idea of grid-dependent thresholds and use a relative vorticity threshold in their work with four different model resolutions between

$\sim 180\,\text{km}$ and $\sim 20\,\text{km}$. In Murakami et al. (2012) a warm core criterion is additionally linked to the resolution.
However, other answers are also found to the problem of resolution dependence: Tory et al. (2013a,b) use the Okubo-Weiss-Zeta (OWZ) parameter, which is independent of the ability of a model to simulate a TC in detail. Thus, the parameter indicates large-scale advantageous conditions for a TC. The OWZ parameter must be $> 50 \cdot 10^{-6}$ at 850 hPa and $> 40 \cdot 10^{-6}$ at 500 hPa. Furthermore, the storms are identified based on a relative humidity (RH) of $> 70\,\%$ and $> 50\,\%$ at 950 and 700 hPa and a wind shear of $< 25\,\text{ms}^{-1}$ between 850 and 200 hPa. Tory et al. (2013a) state that their scheme detects 78 % of the TCs compared to a false alarm rate of 25 %.
Strachan et al. (2013) address the problem of grid dependence differently: they interpolate all data to a resolution of T42 before applying a relaxed relative vorticity criterion. After converting the data to a T63 resolution ($\sim 210\,\text{km}$), a stricter relative vorticity criterion is used to finally identify the TC candidates.
A completely new approach is taken by Liu et al. (2016) who use machine learning via a convolutional neural network (CNN) that is trained on labeled data sets and detects TCs with an accuracy of 99 %. The CNN is applied to the input variables SLP, wind speed at the surface and in 850 hPa, temperatures at 200 and 500 hPa and total precipitable water.

A crucial point is the way in which the detectors are tuned and evaluated. If this happens based on models, the error of the detection scheme and the model error to correctly reproduce the TC climatology add up. This error can be reduced by tuning the detector using several models (Tory et al. 2013a). Nevertheless, tuning and evaluation based on reanalyses is preferable, since the detections can be compared with observations. However, reanalyses also contain a certain error and, as mentioned above, do not represent TCs intensively enough due to their restricted resolution.

In summary, there are three different approaches for the detection of TCs:

- Absolute thresholds, which can only be applied reasonably to a certain resolution. This can be preceded by a conversion of the data set to a uniform resolution so that the method can also be used for other resolutions.
- Relative thresholds that vary either regionally or resolution-dependently
- Methods without fixed thresholds using machine learning

3 | Data and methods

In the following section, the required data sets (Sections 3.1 and 3.2), the principles of the MDI algorithm (Section 3.3) and the evaluation methods (Section 3.4) are presented. Furthermore, ideas for post-processing the detections using filters are presented (Section 3.5).

3.1 ERA-Interim Data

To take advantage of the full range of the capabilities of the MDI algorithm, it is applied to spatio-temporal data that are gridded. To be able to verify the detection skill of the MDI algorithm, reanalysis data has to be used. This allows to decide on past events, which instances represent TCs and which a different weather phenomenon.
The ERA-Interim reanalysis is chosen, which aims to reproduce the correct state of the atmosphere between 1979 and 2019 (Dee et al. 2011). Reanalyses link observations with the short-term prediction of a global circulation model (GCM) to provide gridded data at regular temporal intervals on many pressure levels. The ERA-interim reanalysis has a spatial resolution of 256×512 grid points, i.e. in the context of climate models the resolution T255, which results in about 80 km between the grid points. At the equator the distance is about 60 km. In the vertical, the variables are available at 60 levels between the earth's surface and 0.1 hPa. The time resolution is 6 h, so that the data are available at the times 0 UTC, 6 UTC, 12 UTC, 18 UTC.
Reanalyses tend to underestimate the number of TCs (Murakami 2014) and their intensities (Hodges et al. 2017; Schenkel and Hart 2012). However, the spatial and temporal distribution is usually well represented in comparison with the observations.
Since the MDI algorithm can be applied not only univariately to data sets of individual atmospheric variables, but also multivariately, there are many possible combinations if the selection of variables is not restricted. With about 25 different variables at different pressure levels, which one can extract from Section 2.3, there would be more than 33.5 million possible combinations:

$$\sum_{k=1}^{25} \begin{pmatrix} 25 \\ k \end{pmatrix} = \begin{pmatrix} 25 \\ 1 \end{pmatrix} + \begin{pmatrix} 25 \\ 2 \end{pmatrix} + \cdots + \begin{pmatrix} 25 \\ 25 \end{pmatrix} = 2^{25} - 1 \tag{3.1}$$

Note that here a variable at two different pressure levels, for example relative vorticity at 850 and 250 hPa, is interpreted as two different variables. As a first step, a limited number of important variables must be identified that are noticeably influenced by a TC passage. Furthermore, one has to make sure that the variables have an approximately

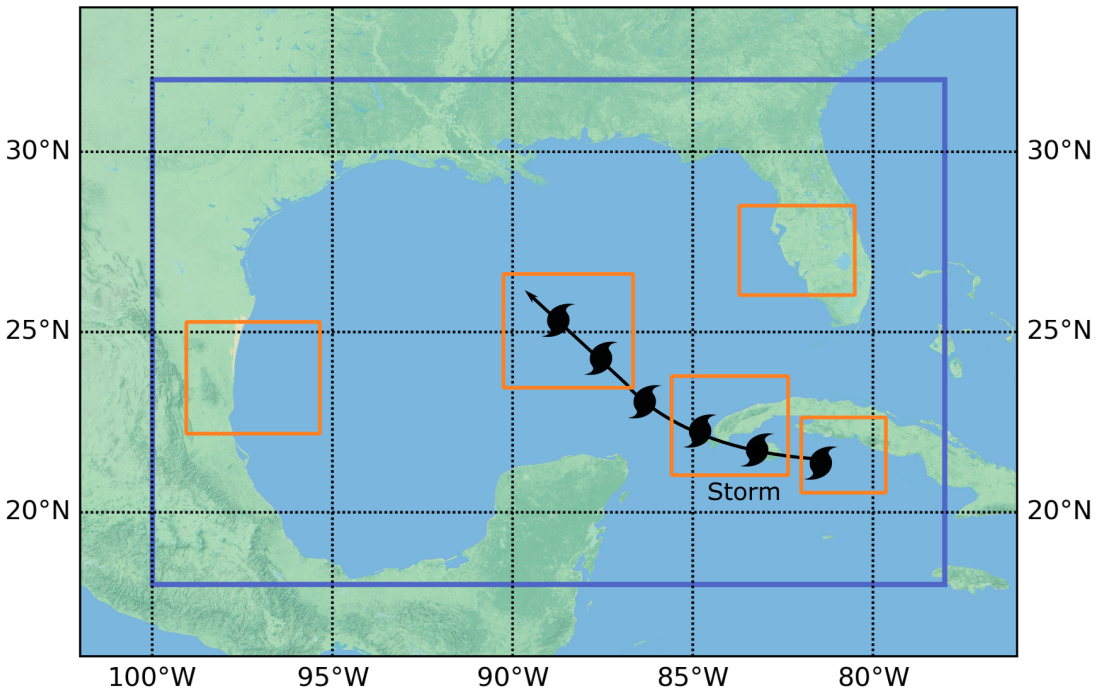


FIGURE 3.1: The blue box shows the analyzed area of ERA-Interim data spanning 260 to 282°E and 18 to 32°N. The black storm track and orange detection boxes are used to explain how the detections of the MDI algorithm are matched to the IBTrACS storm positions (Section 3.4).

normal distribution, which is not the case, for example, with variables such as cloud cover. This will be done by means of a case study in which time series of the ERA-Interim data at a grid point near the Bahamas are examined.

Later, the MDI algorithm is applied to ERA-Interim data spanning the Gulf of Mexico as shown in Figure 3.1 by the blue box. Here, the area covers 260 to 282°E and 18 to 32°N. The hurricane seasons of the years 2000-2010 are analyzed. They start on 1 June and end on 30 November. In order not to miss any early TCs, the period is extended by half a month before and after, resulting in an annually analyzed period from 15 May to 15 December. With the above-mentioned temporal resolution of 6 h, this results in 1460 time steps in total per hurricane season.

As a preprocessing step, the individual variables were normalized. This was done by subtracting their mean and dividing them by their maximum value.

3.2 IBTrACS

To evaluate whether a detection of the MDI algorithm is a TC or not, the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al. 2010) in version v03r10 released in September 2018 is consulted. It provides the most complete set of six-hourly information like position, wind speed and classification of all known historical TCs and subtropical storms worldwide since 1851. Events that occurred outside the analyzed grid box and time, and events that are not classified as tropical storms (TS) with wind speeds < 34 kn are ignored.

Even the combination of ERA-Interim data and the IBTrACS database will contain errors since some circulations may not be present or simulated too weak in the reanalysis data (Murakami 2014) or storms are missing in the IBTrACS database, but since the focus lies on data after the year 2000, this error is considered to be small.

3.3 Maximally Divergent Intervals (MDI) algorithm

The central tool for this thesis is the Maximally Divergent Intervals (MDI) algorithm developed by Barz et al. (2017). In contrast to many existing methods for the detection of anomalies (e.g. Chawla and Sun 2006; Cheng and Li 2006; Hotelling 1947) the MDI algorithm does not scan for pointwise anomalies, but for intervals in space and time that differ from the remaining data series. This is useful for TCs, where not only a single data point is anomalous, but the TC has a certain lifetime and extends over several hundred kilometers. Moreover, only the combined observation of individual points can reveal an anomaly (Barz et al. 2018) if, for example, a remarkable frequency change occurs in the data series.
Not only the application of an unsupervised machine learning algorithm for the detection of TCs is novel, but also the tool itself is pioneering in its unlimited applicability to spatio-temporal data as shown in Section 2.2.

3.3.1 Basic idea

First the theoretical principle of the MDI algorithm is examined: the algorithm is applied to a multivariate, spatial-temporal data set $\mathfrak{X} \in \mathbb{R}^{T \times X \times Y \times Z \times D}$, which is present as a 5-dimensional tensor. Each of its vectors has 4 contextual attributes, three of them spatial and one temporal. The last attribute consists of D behavioral sub-attributes, which are atmospheric variables in this study. A single sample $\mathfrak{X}_i$ is referred to with 4-tuples $i \in \mathbb{N}^4$ to determine the position in space and time.
Now, intervals with the index boundaries $[l, r)$ are formed, which are noted as $\{t \in \mathbb{N} \mid l \leq t < r\}$. All possible intervals with a size between a and b along axis n can be notated as follows:

$$\mathfrak{I}_{a,b}^{n} := \{[l, r) \mid 1 \leq l < r \leq n + 1 \wedge a \leq r - l \leq b\} \tag{3.2}$$

All sub-blocks of the tensor $\mathfrak{X}$ that meet the requirements of the size restrictions $A = (a_t, a_x, a_y, a_z), B = (b_t, b_x, b_y, b_z)$ can be described as follows:

$$\mathfrak{I}_{A,B} := \{I_t \times I_x \times I_y \times I_z \mid I_t \in \mathfrak{I}_{a_t,b_t}^{T} \wedge I_x \in \mathfrak{I}_{a_x,b_x}^{X} \wedge I_y \in \mathfrak{I}_{a_y,b_y}^{Y} \wedge \mathfrak{I}_z \in I_{a_z,b_z}^{Z}\} \tag{3.3}$$

In the following, the indices are neglected and simply $\mathfrak{I}$ is written, since the definition in this context is clear. For each of these sub-blocks $I \in \mathfrak{I}_{A,B}$, one can characterize the

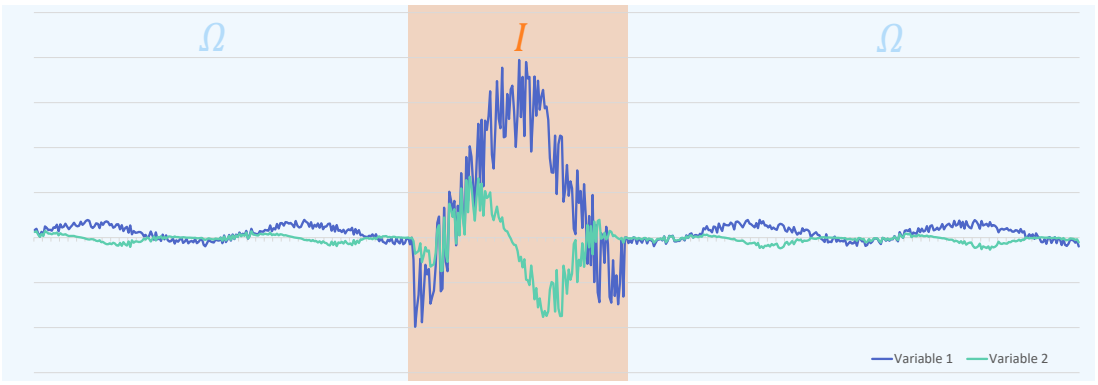


FIGURE 3.2: The basic idea of the MDI algorithm: in a multivariate time series, intervals I (represented in orange) are searched, whose inner variable distribution deviates strongly from the distribution of the remainder Ω (light blue). The idea can also be transferred to spatio-temporal data.

remainder of the data set as:

$$\Omega(I) := ([1, T] \times [1, X] \times [1, Y] \times [1, Z]) \setminus I \tag{3.4}$$

The fundamental idea of the algorithm is to find intervals I that deviate as much as possible from the remaining data set Ω as shown in Figure 3.2. For this, the inner data distribution p_I and the outer data distribution p_Ω are modelled with multivariate Gaussian distributions and then compared.
The comparison is performed using a metric $\mathfrak{D}(p_I, p_\Omega)$, which measures the deviation between the two distributions.
Mathematically, the task of the algorithm can therefore be formulated as finding the intervals with the greatest deviation from the rest of the data set:

$$\hat{I} = \underset{I \in \mathfrak{I}_{A,B}}{\arg\max}\, \mathfrak{D}(p_I, p_{\Omega(I)}) \tag{3.5}$$

For this, the algorithm analyzes all intervals $I \in \mathfrak{I}_{A,B}$, models the distributions p_I and p_Ω, and finally calculates the respective divergence $\mathfrak{D}$. This divergence, hereinafter referred to as the *score* of a detection, is used to rank the intervals in a list, which is returned to the user. Using the previous knowledge of TCs, one can define the interval limits a and b in advance. It is expected that this leads to a better detection skill in TC detection: a spatial interval size of 3 to 8 grid distances is defined, both zonal and meridional. In case of the resolution of the used ERA-Interim data, an approximate spatial extension of between $\sim 225\,\text{km}$ and $\sim 800\,\text{km}$ is considered. This appears to be a reasonable assumption, as most TCs in the North Atlantic and North Pacific have an extension between 150 km and 500 km in terms of their outermost closed isobar (Merrill 1984). Since a TC moves during the time that a detection box is present, an area such as the aforementioned is influenced. An even lower limit is unfavorable because the estimation of the inner distribution becomes inaccurate due to the small number of included data points. Larger boxes contradict the dynamical motion of a TC. Large, immobile detection boxes can also be traversed diagonally, so that only a small part

of the detection box shows an anomalous distribution of the variables. This lowers the score which this box receives. Medium-sized detection boxes, which will at best be placed along the storm path like outlined in Figure 3.1, sound more promising. The time extension for one detection box is limited to 2 to 8 time steps, i.e. 12 to 48 h.

3.3.2 Modelled distributions

The MDI algorithm provides two ways to model the p_I and p_Ω distributions that are required to calculate the divergence metric $\mathfrak{D}$ according to Equation 3.5. The algorithm calculates these distributions for each interval that is suitable in general, so its calculation must be effective. The first model is the kernel density estimation (KDE), which, however, is difficult to apply to long time series and does not consider correlation between attributes (Barz et al. 2018). In the second model, one assumes that both the data in interval I and the remainder Ω obey a multivariate Gaussian distribution. Thus, they are modelled by $\mathcal{N}(\mu_I, S_I)$ and $\mathcal{N}(\mu_\Omega, S_\Omega)$, where $\mu \in \mathbb{R}$ is the expected value and S the standard deviation. Since the MDI algorithm will be applied to a large-scale data set and a fast procedure is required, and since the correlations between the variables could also be of importance in the detection of hurricanes, the Gaussian distribution model is chosen. The computational complexity of the MDI algorithm in this case is $\mathcal{O}(N \cdot L^2)$, where $N := T \cdot X \cdot Y \cdot Z$ is the number of samples and $L = (b_t - a_t + 1) \cdot (b_x - a_x + 1) \cdot (b_y - a_y + 1) \cdot (b_z - a_z + 1)$ denotes the maximum volume of an interval.

3.3.3 Divergences

It is necessary to find a metric $\mathfrak{D}$ to calculate how much the distribution p_I deviates from the remainder of the data with the distribution p_Ω. There are several possibilities to choose from, which are already implemented in the MDI algorithm and will be tested in this thesis for their suitability to detect TCs.
The methods are all derived from information theory and are based on the concept of the entropy that was introduced by Shannon (1948):

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{3.6}$$

where P is a set of events that have the probability of occurrence $p_1, p_1, ..., p_n$. In this sense, entropy can be interpreted as the uncertainty of an experiment. With an equally distributed probability $p_1 = p_2 = p_n = \frac{1}{n}$, as it is given in a fair dice experiment, it reaches its maximum, $\log(n)$. In this type of experiment, the highest uncertainty exists regarding the result. The more uncertain the outcome of an experiment is, the more information can be obtained from it. With an experiment that allows two different events with the probabilities $p_1 = 1$ and $p_2 = 0$, one does not gain any information when performing it, because the result is clear beforehand. Therefore, entropy can also be seen as the average information content per realization of an experiment.

The extended concept of cross entropy

$$\mathfrak{D}_{CE}(p,q) := H(p,q) := \mathbb{E}_p[-\log\ q] \tag{3.7}$$

is a measure of how surprising a sample of p is, assuming it would have been drawn from q. $\mathbb{E}_p$ is the expected value based on the distribution p. The more different the distributions p and q are, the more unexpected it is to obtain a sample that reflects the distribution q, even though it was drawn from distribution p. The entropy increases and is therefore a suitable measure to quantify the divergence of two intervals. The cross entropy can be approximated as

$$\widetilde{\mathfrak{D}}_{CE}(I,\Omega) = \frac{1}{|I|}\sum_{i\in I}\log p_\Omega(\mathfrak{X}_i). \tag{3.8}$$

It is advantageous that only the distribution p_Ω has to be approximated. The distribution p_I, which usually contains fewer data points and is as such difficult to model, is not required for the calculation.

The Kullback-Leibler (KL) divergence is a further enhancement of the cross entropy and has the advantage that it additionally incorporates the intrinsic entropy of the distribution p:

$$\mathfrak{D}_{KL}(p,q) := H(p,q) - H(p,p) = \mathbb{E}\left[-\log\frac{p}{q}\right] \tag{3.9}$$

The advantage of considering the intrinsic entropy $H(p,p) := H(p)$ of the examined interval is that an interval with a stable distribution receives a higher score than a strongly fluctuating interval, assuming that both intervals have the same cross entropy with the remainder of the data. Note that $H(p)$ has a negative sign (see Equation 3.6), which is why $\mathfrak{D}_{KL}(p,q)$ becomes larger.

As with cross entropy before, one can estimate the KL divergence using discrete values from the data series:

$$\widetilde{\mathfrak{D}}_{KL}(I,\Omega) = \frac{1}{|I|}\sum_{i\in I}\log\left(\frac{p_I(\mathfrak{X}_i)}{p_\Omega(\mathfrak{X}_i)}\right) \tag{3.10}$$

In contrast to cross entropy, also the p_I distribution has to estimated, which makes the algorithm more complex and time-consuming.

A special feature of the MDI algorithm is that it allows the comparison of intervals of different sizes. This leads to another problem of the KL divergence: shorter time intervals receive a higher score, which leads to the consequence that a temporally extended anomaly is divided into many small detections. Barz et al. (2018) therefore introduced the unbiased Kullback-Leibler (U-KL) divergence, which is weighted by the length of the interval $|I|$:

$$\mathfrak{D}_{\mathrm{U-KL}}\left(p_I, p_\Omega\right) := 2\cdot|I|\cdot\mathfrak{D}_{\mathrm{KL}}\left(p_I, p_\Omega\right) \tag{3.11}$$

Barz et al. (2018) demonstrated that the U-KL divergence is suitable for anomaly de-

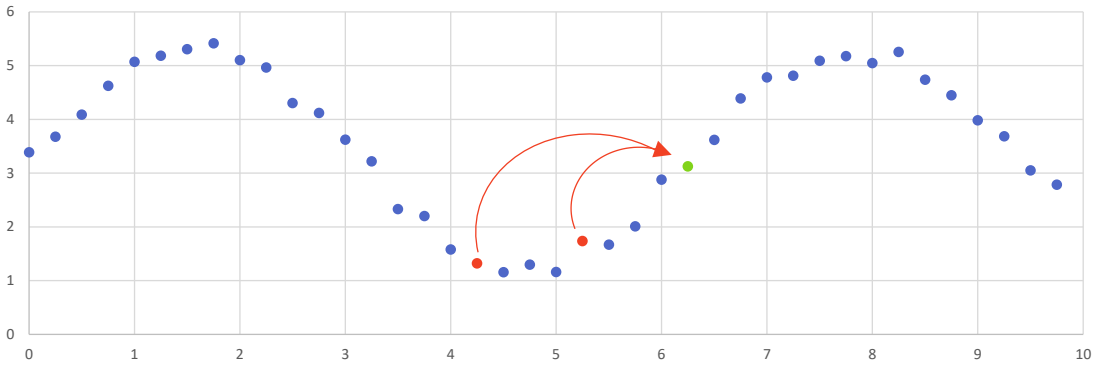


FIGURE 3.3: Principle of time-delay embedding: the green data point is replaced by a $\kappa = 3$-dimensional vector, which additionally incorporates the red marked data points, located in this example in a distance of $\tau = 4$ time steps. Reproduced after Barz et al. (2017).

tection in finitely long time series, i.e. it can also be used in the case of TC detection.

3.3.4 Embedding

The models used to estimate probability density of the intervals assume that the samples are independent of each other. In the case of a climatological time series, this is not the case because neighboring data points influence each other and there is a correlation between them. One approach to address this circumstance is the preprocessing step of *embedding*, in which the individual data points of a data set are replaced by vectors, into which additionally surrounding data points are incorporated. These data points have a certain distance to the actual data point.

The so-called *time-delay embedding* (Packard et al. 1980) transforms a given time series $(x_t)_{t=1}^n$, $x_t \in \mathbb{R}^D$ into a time series $(x'_t)_{t=1+(\kappa-1)\cdot\tau}^n$, $x'_t \in \mathbb{R}^{\kappa D}$, including the elements

$$x'_t = \left(x_t^\top \quad x_{t-\tau}^\top \quad x_{t-2\tau}^\top \quad \cdots \quad x_{t-(\kappa-1)\cdot\tau}^\top\right)^\top \tag{3.12}$$

where κ represents the dimension of the resulting vectors, i.e. how many data points at a point are combined to form a vector. τ is the temporal step by which the points are selected. An example for illustration is shown in Figure 3.3.

The basic principle behind this preprocessing step is Takens' theorem (Takens 1981), according to which the status of a chaotic dynamic system can be reconstructed by considering κ observations of a dynamic system. One of the tasks of this master thesis will be the selection of the most suitable parameter τ and κ to optimize the MDI algorithm for TC detection.

Not only temporally but also spatially context can be incorporated in the investigated intervals. This is done in *spatial-neighbor embedding*, where the three parameters κ_x, κ_y and κ_z have to be defined. As before, they specify the number of adjacent values that are merged into a vector at any point – in this case not temporally, but spatially neighboring points. The spacing between the included grid points is specified by τ_x, τ_y and τ_z. Examples are visually illustrated in Figure 3.4. Unlike time-delay embedding, not only the previous, but also the grid points left and right along the x-axis, above

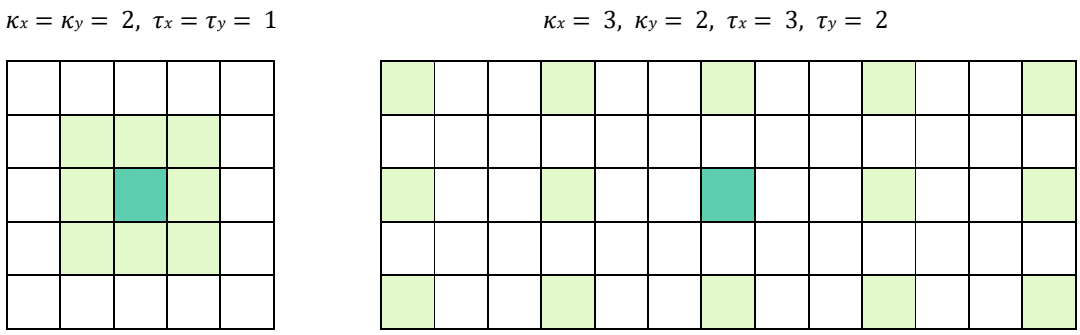


FIGURE 3.4: Two examples to illustrate spatial-neighbor embedding: instead of the turquoise middle data point, the surrounding light green data points are also incorporated to create a higher dimensional vector. Reproduced after Barz et al. (2017).

TABLE 3.1: Schematic of the contingency table for verification of categorical data.

	Observed event	No observed event
Detection	Hit	False alarm
No detection	Miss	Correct rejection

and below along the y-axis and lower and higher along the z-axis are included. Thus, when using spatial-neighbor embedding, the amount of data to be considered increases substantially.

3.4 Detection skill

Any detection box that contains a ground truth for $\geq 25\,\%$ of its present time steps is counted as correct. Each ground truth contained therein is counted as a hit (H). If this is not the case or a detection box does not contain any ground truth, it is classified as false alarm (F). Ground truths that are not enclosed by a detection box are labeled as miss (M). Table 3.1 gives an overview of the terms.
Furthermore, some categorical statistical measures (Wilks 2011) are introduced and computed from the contingency table. The probability of detection (POD) is defined as

$$\mathrm{POD} = \frac{H}{H + M}. \tag{3.13}$$

This measure is the total number of hits divided by the total number of observed TCs and gives the probability by which an observed TC is detected. The POD is sometimes also referred to as *recall*. An ideal detector would yield a POD of 1.
The false alarm rate (FAR) measures how many of the found TCs are falsely detected. It is defined as follows:

$$\mathrm{FAR} = \frac{F}{H + F} \tag{3.14}$$

In the ideal case a detector achieves $\mathrm{FAR} = 0$.

The precision (PRC) is calculated as

$$\mathrm{PRC} = \frac{H}{H+F} \tag{3.15}$$

and indicates how precise a detector is when searching for a certain class of objects, i.e. how many of all its detections ($H + F$) are hits. If FAR = 0, then PRC = 1.
Another measure is the bias score (BIAS) which compares the total number of detected TCs and the total number of observed TCs:

$$\mathrm{BIAS} = \frac{H+F}{H+M}. \tag{3.16}$$

A BIAS of 1.0 would mean that the detection scheme detects as many TCs as there are in the observational data set, independent of the fact that the detections are real hits.

In the case of the MDI algorithm, a problem is that it returns up to 100000 detections. Among them there are only a few hits and many false alarms. FAR and PRC would indicate poor performance of the algorithm. Moreover, the list of detections cannot yet be stopped at a certain point, because the number of true detections varies depending on the application. Therefore, a metric is needed for which it is relevant that the hits are placed at the top of the detection list (i.e. that they receive high scores). The many false alarms on the lower ranks of the list should be irrelevant. These requirements are met by the Average Precision (AP): for every possible detection cutoff rank n, the PRC and POD (*recall*) are calculated. The Average Precision (AP), which is independent of a specific cutoff rank n, is then calculated as

$$AP = \int_0^1 p(r)\, dr \tag{3.17}$$

where p is the continuous function of PRC and r is for recall (POD). This measure is adopted from the field of information retrieval, where the aim is to return as many relevant (text) documents as possible to a user in response to his request, whereby the relevant documents should be listed as high up in the ranking as possible. If one displays the PRC and recall value pairs as a precision-recall curve in a diagram as shown in Figure 3.5, the area below the curve represents the AP. In the case of discrete values, the AP is calculated as

$$AP = \sum_{n=1}^{N} \mathrm{PRC}(n)\, \Delta r(n). \tag{3.18}$$

PRC(n) represents the PRC at a specific cutoff rank n of detections. N is the total number of detections and $\Delta r(n)$ is for the change of recall between the current and the previous cutoff rank $n-1$. In the optimal case, the AP is 1.0. In this situation all true detections are present in the upper part of the ranking, are not interrupted by false alarms and all ground truths are detected. AP is particularly useful when analyzing data sets that have an imbalanced class distribution, i.e. where there are less positive than negative instances, or vice versa (Saito and Rehmsmeier 2015). This is the case in

TABLE 3.2: Exemplary ranking of detections in three different cases: an ideal case where the first five detections are all hits, an adverse case where the five true detections are in 6 to 10^{th} rank and a realistic case where the order is mixed. For the realistic case, the number of hits, false alarms and misses, as well as the corresponding recall and PRC, are also displayed if only the first n detections in the ranking are considered. For the calculation of the metrics a total number of 7 ground truths was assumed in the data set, which could have been detected.

Rank	Ideal case	Adverse case	Realistic case	Hits	False alarms	Misses	POD (recall)	PRC
1	H	F	H	1	0	6	0.14	1.00
2	H	F	H	2	0	5	0.29	1.00
3	H	F	F	2	1	5	0.29	0.67
4	H	F	H	3	1	4	0.43	0.75
5	H	F	F	3	2	4	0.43	0.60
6	F	H	H	4	2	3	0.57	0.67
7	F	H	F	4	3	3	0.57	0.57
8	F	H	F	4	4	3	0.57	0.50
9	F	H	H	5	4	2	0.71	0.56
10	F	H	F	5	5	2	0.71	0.50

the used data set, as TC only occur in a few places compared to the total number of possible detected intervals.

In Table 3.2, three examples are illustrated. In the *ideal case*, there are many hits among the top detections, as shown in column 1. The false alarms, on the other hand, are on the lower ranks. The opposite is shown in the next column as the *adverse case*: here the false alarms receive a high ranking, whereas the true detections are placed on the lower ranks. In reality, there will be a mixture of these cases, as shown in the *realistic case*. For this case, hits, false alarms, misses, POD (recall) and PRC are also shown when considering the first n detections. To calculate these metrics, a total of 7 ground truths is assumed, which the algorithm could at best detect.

The more detections of the returned list are included, the higher the POD (recall) mostly will be. However, the PRC will also decrease in this manner, as some detections will be false alarms. In Figure 3.5 the curves for the rankings of Table 3.2 are shown. In the ideal case, shown in green, not all ground truths are detected, but the algorithm recognizes this object class well. The PRC only drops at the end when more than the first 5 detections are considered. Already visually one can see that the area under the green curve is larger than the areas under the other curves and the AP is therefore the highest.

The AP can be calculated for the detection of various object classes q, in meteorological applications for example different weather phenomena (TCs, droughts, atmospheric rivers). If a mean of different APs is formed, this is referred to as the mean Average Precision (mAP)

$$\text{mAP} = \frac{\sum_{q=1}^{Q} \text{AP}(q)}{Q} \tag{3.19}$$

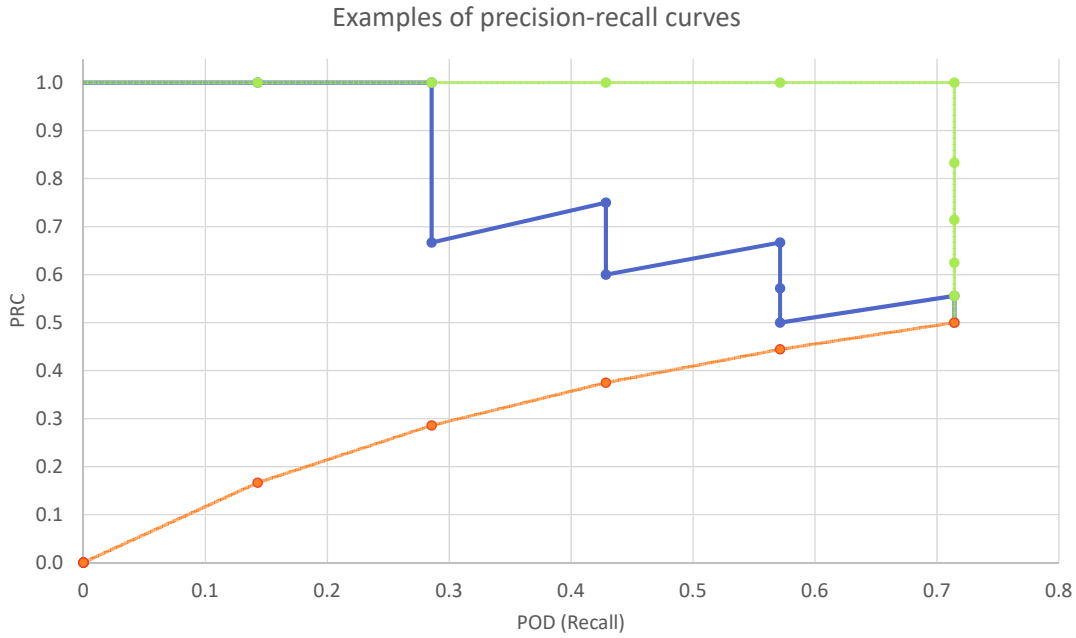


FIGURE 3.5: Exemplary precision-recall curves, as they would result from Table 3.2. The value pairs for the realistic case are explicitly given in that table. For a meaningful calculation of the AP, i.e. the area under the curves, a horizontal line is drawn between the first pair of values and the y-axis (unless this data point is already at $POD = 0$).

where Q describes the number of all classes one searches for. This term will also be used if the average over multiple APs, such as the APs of different hurricane seasons is formed.

3.5 Filtering of detections

In Section 3.4 it was mentioned that the MDI algorithm returns a list of many detections and that it is not possible to find an exact cutoff rank in advance. Therefore, other methods to filter the detections are necessary. For this purpose, one could use soft limits, which do not contradict the original idea of developing a TC detector that operates without any hard thresholds. The main aim of the filters will be to remove calm detections. Indeed, the MDI algorithm does not provide any information on why it considers an interval to be anomalous. For example, particularly calm, steady intervals, such as droughts, are just as much anomalies as turbulent phases, represented by storms. Therefore, four different filtering methods have been developed:

(1) Wind speed filter: a straightforward approach is to exclude events that show hardly any or low wind speeds at the surface layer. In the code, this is realized by only preserving detections that have a certain wind speed at least at one of their grid points. Furthermore, this condition must be fulfilled during $\geq 50\,\%$ of the time steps covered by the detection box. The exact value is varied and it is observed how this change affects the detection skill of the MDI algorithm in TC detection.

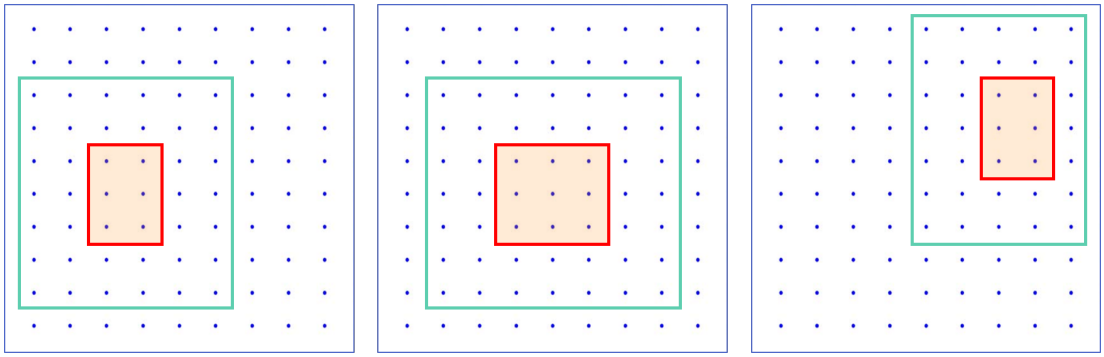


FIGURE 3.6: Illustration of the low-pressure filter. Three exemplary detection boxes are shown in red. They are surrounded by a turquoise comparison box. To be retained, the average geopotential height at 1000 hPa of a detection box must be below the one of the respective turquoise box during $\geq 50\,\%$ of the time steps that the detection box is present. If a detection is located close to the edge of the examined area (blue border), the size of the comparison box is reduced accordingly on this side.

(2) Relative vorticity filter: the procedure is similar to the first filter, but now using the relative vorticity at 850 hPa. As an additional condition, the averaged relative vorticity of the detection box must be positive for $\geq 50\,\%$ of the time steps to ensure a cyclonic system.

(3) Low-pressure filter: this filter is designed to filter out detections that are not low-pressure areas. For this purpose, a second box is placed around the detection box, which is larger by a certain number of grid distances. In Figure 3.6 this is shown exemplary with two grid spacings. During $\geq 50\,\%$ of the time steps, the average geopotential height at 1000 hPa of the detection box must be lower than in the surrounding box. How many grid spacings the comparison box is larger is varied in a similar way as with the previous filters.

(4) Cutoff threshold: since the MDI algorithm provides its detections with scores, a score threshold can be defined below which the detections are discarded. The exact value is not trivial and is elaborated using the precision-recall curves. The value is chosen so that a PRC of at least 0.8 is achieved.

4 | Results

Now that the methods and data sets that are essential for this study have been addressed, the results of the initial case study to limit the variables to which the MDI algorithm should be applied are presented (Section 4.1). Thereafter, it is analyzed which initial settings should be selected (Section 4.2), to which combination of variables the MDI algorithm should be applied (Section 4.3) and how the detections should be post-processed (Section 4.4).

4.1 Case study to select the input variables

As mentioned in Section 3.1, the number of variables described in Section 2.3 needs to be restricted, as otherwise there are many possible combinations of variables due to the multivariate applicability of the MDI algorithm. Therefore, a case study was conducted first to identify variables that already visually show anomalies in a time series if the corresponding grid point is influenced by a TC.
To do this, a location near the Bahamas (24.0 N, 68.25 W) like in Barz et al. (2017) is chosen and the time series of 16 variables are analyzed which are available in the ERA-Interim data set. The respective variables show an approximate normal distribution over the period considered. Consequently, the MDI could be applied to them (see Section 3.3.2). The time period is the North Atlantic hurricane seasons 2000-2011. Figure 4.1 shows exemplary the time series of the 16 variables for the season 2004. Time periods during which a TC with a wind speed of at least 34 kn, i.e. a tropical storm (TS), has passed the above-mentioned coordinates within five grid cells are shaded in light blue. In the example of the 2004 hurricane season, two events of this kind are visible: the first is hurricane Frances, which hit the Bahamas in early September, the second is hurricane Jeanne, which occurred after mid-September.
Through the visual analysis of all hurricane seasons from 2000 to 2010, the following variables were identified as the most promising: the geopotential height at 1000 hPa (z1000), temperature at 500 hPa (t500), wind speed at 300 hPa (wind300) relative vorticity at 250 and 850 hPa (vo250 and vo850) as well as absolute humidity at 700 hPa (q700). The relative vorticity is taken into account on two different pressure levels as it is considered to be particularly important for the detection of TCs. Zarzycki and Ullrich (2017) suggest the geopotential height difference between 300 and 500 hPa (dz300500), as a larger layer thickness is typical for a warm core of a hurricane. Alternatively, they consider temperature at 400 hPa (t400) as a reasonable variable, which is why t500 is replaced by this variable in this study. As a result, the number of variables which the

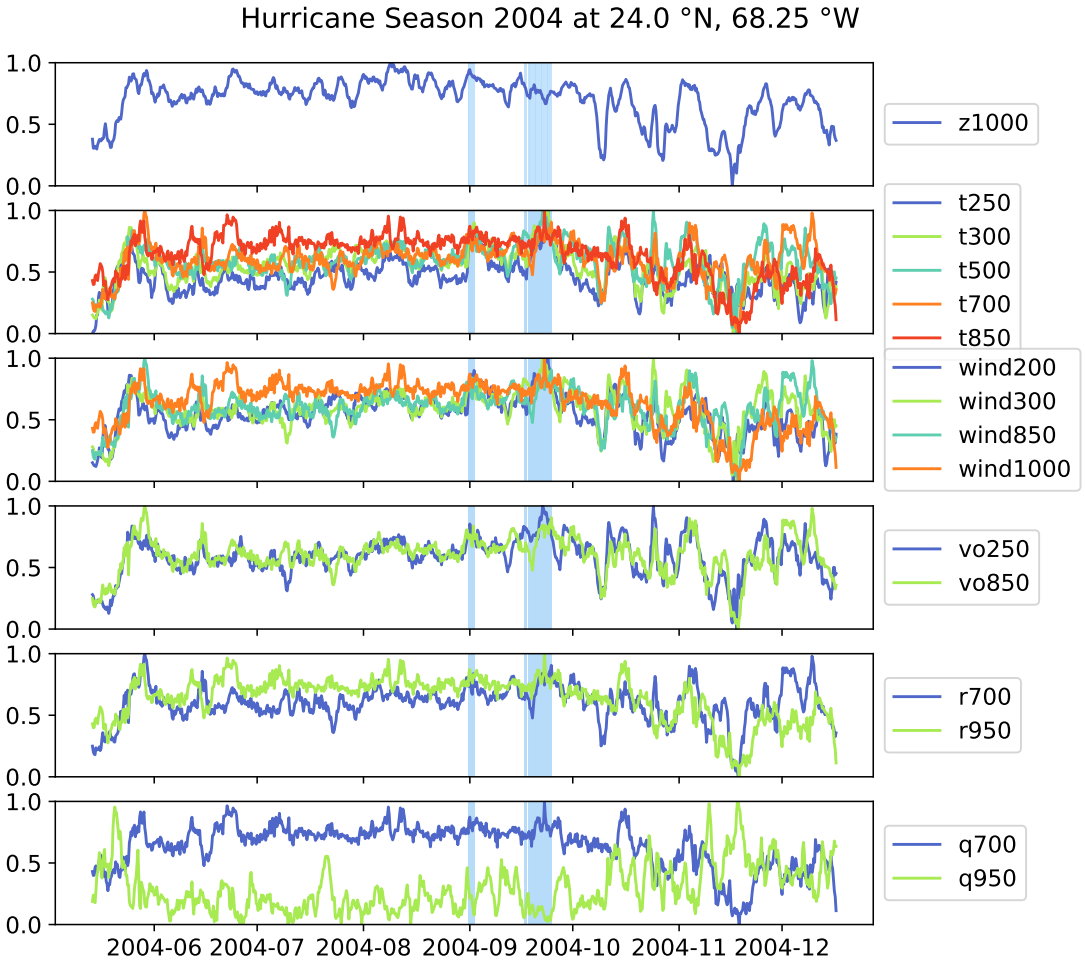


FIGURE 4.1: Figure of the case study conducted to determine variables most affected by a TC passage. The 2004 hurricane season is shown. The time series are extracted from ERA-Interim data at the grid point closest to coordinates 23.866°N, 68.481°W. Periods in which a storm has passed at a distance of five grid points or less are highlighted in blue. Data of actual storms are taken from the IBTrACS data set (see Section 3.2). The numbers behind the different variables indicate the respective pressure levels in hPa. All time series are normalized by subtracting their mean and dividing by their maximum value.

MDI algorithm is applied to in the following has been reduced to 7.

4.2 Optimal initial parameters

In this study, the MDI algorithm was applied to the selection of variables mentioned above. It can be applied to individual variables or to multivariate combinations of variables. For example, the MDI algorithm can be tested on z1000, t500, etc. individually or to all 7 variables simultaneously, but also to all possible combinations of 2, 3, etc. variables. This results in $2^7 - 1 = 127$ different possible combinations (Equation 3.1). The first goal is to find the optimal initial settings of the MDI algorithm. In principle, the following settings and parameters can be varied, which directly influence the behavior of the MDI algorithm:

(1) Divergence metric for calculating the score: cross entropy, KL divergence, U-KL

divergence

(2) Minimum and maximum size of the detection boxes, both spatial (x_{min}, y_{min}, x_{max}, y_{max}) and temporal (t_{min}, t_{max})

(3) Parameters of time-delay embedding (κ and τ). The chosen basic configuration is $\kappa = 3$, $\tau = 3$.

(4) Parameters of spatial-neighbor embedding ($\kappa_x, \kappa_y, \tau_x, \tau_y$). The chosen basic configuration is $\kappa_x = \kappa_y = 1$, $\tau_x = \tau_y = 1$.

Theoretically, there is a wide range of possible combinations of settings, which cannot all be tested due to limited computation time. Therefore, the selection of the ideal divergence method (see Section 3.3.3) is the first step and subsequently the choice of the other parameters mentioned in the bullet points (2) to (4). The size of the detection boxes will not be varied since there are many possible configurations. As described in Section 3.2, the following restrictions on size are considered to be useful:

- $t_{min} = 2$ (12 h), $t_{max} = 8$ (48 h)
- $x_{min} = y_{min} = 3$ (~ 225 km), $x_{max} = y_{max} = 8$ (~ 600 km)

The computational effort to test the effect of parameter changes on all 127 variable combinations in the hurricane seasons 2000-2010 would be high. This would result in $127 \cdot 11 = 1397$ runs per parameter setting. Therefore, 100 randomly selected variable-hurricane seasons combinations have been selected as samples. The same random combinations have been chosen for each changed setting (such as a change of the divergence metric), giving an experiment with repeated measurements. For example, the MDI algorithm is applied to the combination z1000-wind300-vo250-vo850-q700 in the 2006 hurricane season, then to t400-dz300500 in the 2009 hurricane season, etc. until 100 such runs are performed. By analyzing the *ensembles* of 100 samples once using cross entropy and once using the KL divergence, it is possible to compare which of the divergence methods is the preferable one for TC detection.
To calculate the APs, all events with a minimum strength of a TS, i.e. with wind speeds of 34 kn or more, in the IBTrACS database were assumed to be ground truths. This practice is in line with the studies published so far in the field of TC detection.

4.2.1 Determination of the preferable divergence method

With the above-mentioned size of the detection boxes and the basic configuration of the spatial-neighbor embedding parameters, which are all set to 1 here (i.e. no spatial-neighbor embedding takes place), and time-delay embedding parameters, which are set to 3, the optimal divergence method for TC detection was determined. Figure 4.2 shows the results for the three tested divergence methods. When comparing the results, mainly the mean value of the APs is consulted and not the median. This is because the distributions of the ensembles generally appear to be asymmetric, as visible in Figure 4.2. The median is always below the mean value. Upwards, as can be seen from

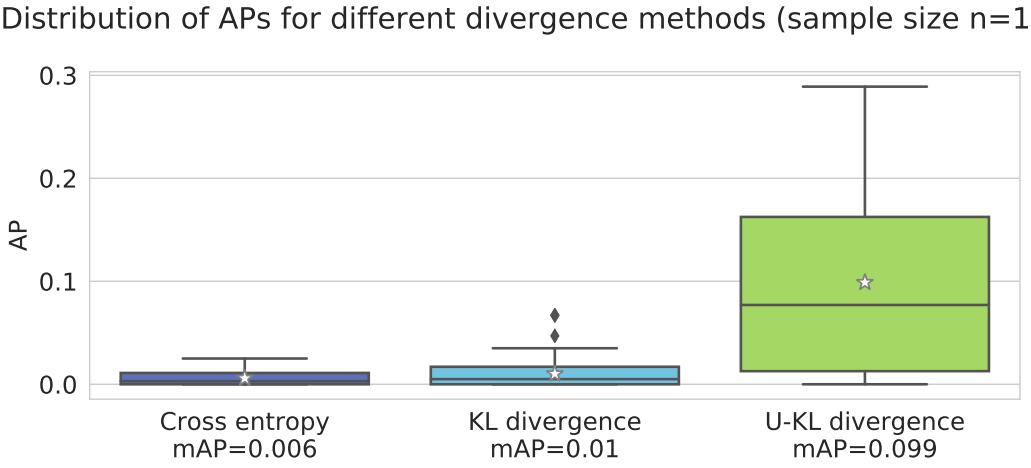


FIGURE 4.2: Box plot showing the distribution of APs using different divergence methods. The MDI algorithm was applied to 100 different variable-hurricane season combinations per divergence method. The boxes show the 25^{th} and 75^{th} percentiles, the medians are displayed as lines inside the boxes, the mean values are marked with white stars. The whiskers are limited to 1.5 times the length of the interquartile range (IQR), but end at the most extreme value within that distance. Values outside 1.5·IQR are shown as black diamonds. Below the boxes, the respective mAPs, i.e. the positions of the stars, are written next to the x-axis.

the long upper whiskers, there are APs that lie far outside the range spanned by the 1^{st} and 3^{rd} quartiles. As the goal is to find the optimal combination of algorithm settings and used variables, the focus is mainly on the high APs and one has to worry less about the overall distribution and the poorly performing variable combinations. This idea is better represented by the mean because it is less stable against the extreme values (to the top).

The cross entropy achieves a mAP of 0.006, the KL divergence 0.01 and the U-KL divergence 0.099. To check whether these results differ statistically significantly, the Friedman test (Friedman 1937, 1940) is used. This is a statistical test with the null hypothesis (H0) that the samples of repeated measurements were taken from the same distribution. The repeated measurements are the ensembles of runs with different divergence methods. The test does not require the sample data to be normally distributed. This test is the non-parametric equivalent to an analysis of variance (ANOVA). If the probability value $p < 0.05$, H0 is rejected and a post-hoc test can be performed. For this, the Nemenyi test (Nemenyi 1963) is conducted, which compares the different ensembles among each other (Demšar 2006). Again, p-values are returned for each pairwise comparison. If $p < 0.05$, the two compared sample ensembles differ and one can speak of a statistically significant improvement or decrease in the detection skill of TCs.

When comparing the ensembles of different divergences, the Friedman test returns $p = 7.854 \cdot 10^{-36}$, which allows to reject H0. The Nemenyi test provides the p-values presented in Table 4.1.

The low p-value of 0.01 in the column of U-KL divergence indicates that significantly better results are obtained by using this divergence compared to the other two methods. Based on the results, the U-KL divergence is chosen for further development of the TC detector.

TABLE 4.1: p-values of the Nemenyi-Friedman test, which compares the AP distributions obtained with the different divergence methods pairwise. The column of the ultimately preferred method, the U-KL divergence, is highlighted by a frame.

	Cross entropy	KL divergence	U-KL divergence
Cross entropy		0.013	0.001
KL divergence	0.013		0.001
U-KL divergence	0.001	0.001	

4.2.2 Determination of the time-delay embedding parameters

The search for the optimal parameters of the time-delay embedding is carried out in a similar way to the comparison of the divergence methods. There are, however, many more ensembles to compare. The following ordered pairs are used for κ and τ: $\{\kappa \in \mathbb{N} \mid 1 \leq \kappa \leq 6\} \times \{\tau \in \mathbb{N} \mid 1 \leq \tau \leq 6\}$. In principle, κ could be further increased, but the computational effort becomes high. With a $\tau > 6$ very distant values would be incorporated in the analyzed interval (see Figure 3.4). In order to assess the maximum range of τ, values between 1 and 6 are used, as described above. In theory, there are 36 combinations to test. However, the pairs (κ, τ)=(1,1), (1,2), (1,3), etc. are identical, since $\kappa = 1$ does not include any surrounding values, no matter how large τ becomes (see Figure 3.3). Effectively, 36 - 5 = 31 runs are performed. Other parameters, as for example for the spatial-neighbor embedding, are not changed yet.
The results are shown as box plots in Figure 4.3. Please note that for the visualization of the first column, only the ensemble of $\kappa = 1$, $\tau = 1$ was calculated and copied to the next rows, as these are actually the same combinations. With a mAP of 0.194, the combination $\kappa = 1$, $\tau = 1$ performs best, $\kappa = 6$, $\tau = 2$ with a mAP value of 0.016 performs worst. Basically, any kind of time-delay embedding reduces the detection skill of the MDI algorithm in detecting TCs. At the same time, time-delay embedding also increases the computational complexity.
The result of $\kappa = 1$, $\tau = 1$ differs significantly from most of the other parameter combinations in the Nemenyi test (see Appendix, Table A.1). Only from the combinations (2,1), (3,1), (4,1), (2,2), (3,2), (2,3) and (2,4) it does not deviate significantly. Based on these results, $\kappa = 1$, $\tau = 1$ is chosen, especially due to the better mean value and its low computational effort.

4.2.3 Determination of the spatial-neighbor embedding parameters

After the finding that every temporal embedding reduces the detection skill of the MDI algorithm in detecting TCs, the focus lays now on spatial-neighbor embedding. Here κ_x, κ_y, τ_x and τ_y will be varied. As mentioned in Section 3.3.4, spatial-neighbor embedding means a higher computational effort, because embedding takes place in both x- and y-direction, as shown in Figure 3.4. Therefore, κ_x and κ_y will be increased to a maximum of 2. In addition, a symmetrical embedding is performed, i.e. $\kappa_x = \kappa_y = \kappa_{xy}$ and $\tau_x = \tau_y = \tau_{xy}$.
τ_{xy} can be increased without additional computational effort, so it will be increased up

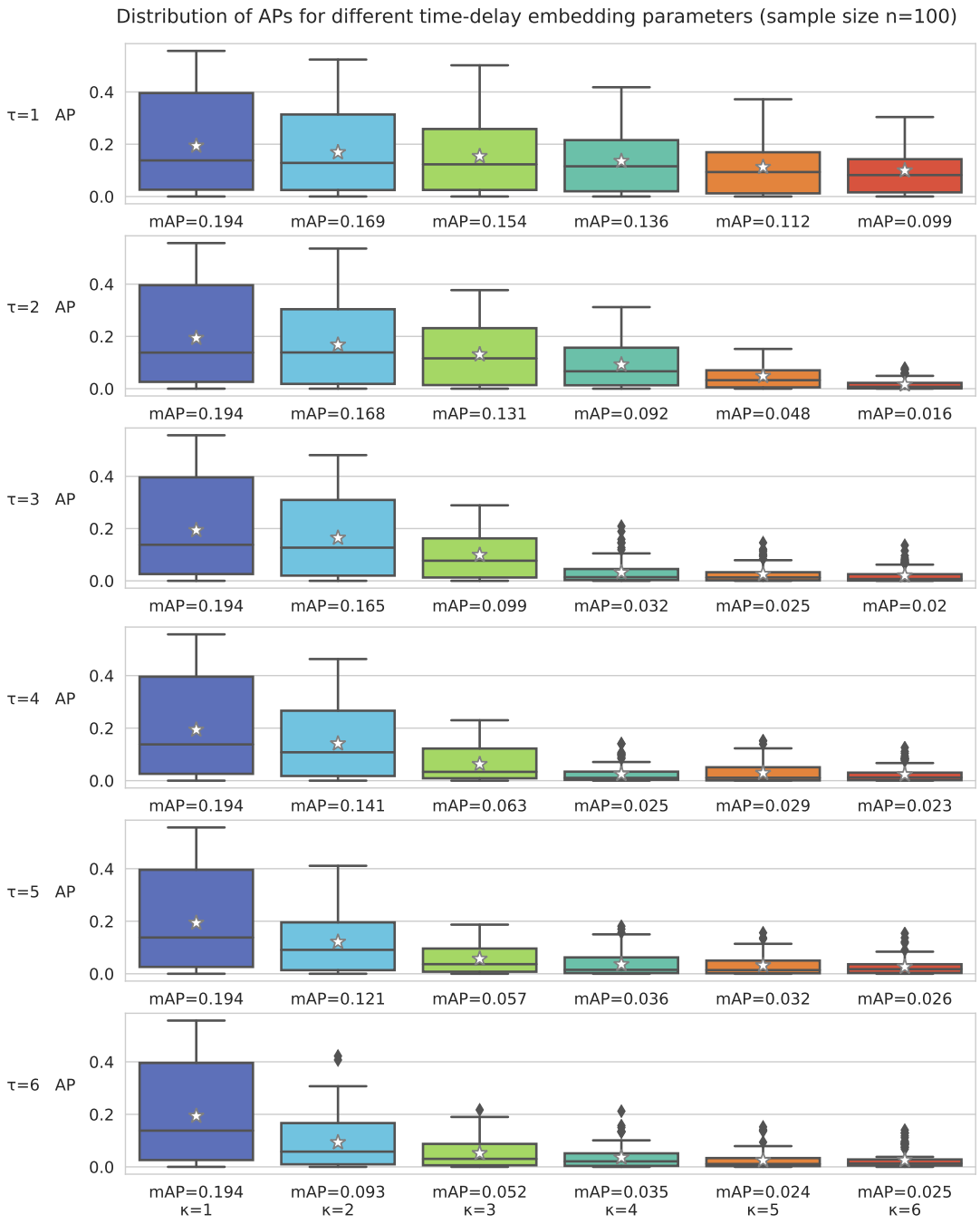


FIGURE 4.3: Box plots showing the distributions of APs using different time-delay parameter combinations. Towards the right, the parameter κ increases (values given below each box plot), from top to bottom, the parameter τ increases.

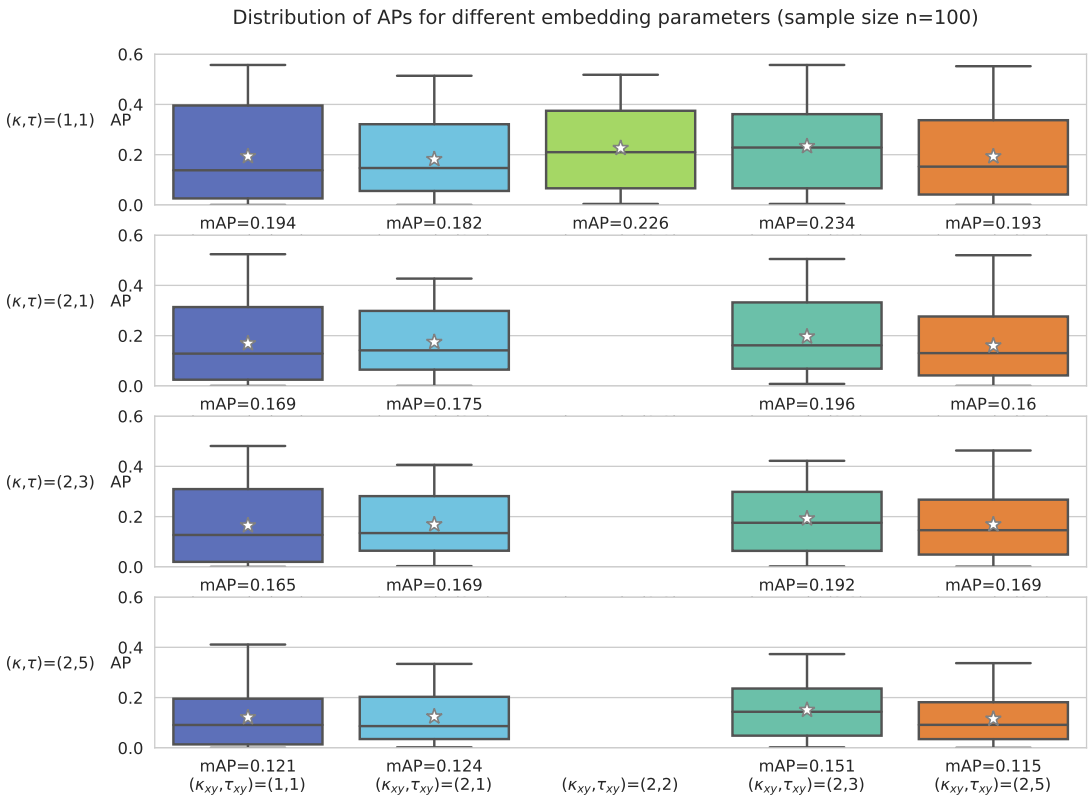


FIGURE 4.4: Box plots, which show the distributions of APs using different embedding parameter combinations. In the vertical direction the size of the time-delay embedding parameters κ and τ varies (noted on the left), in the horizontal direction the size of the spatial parameters κ_{xy} and τ_{xy} (noted at the bottom).

to 6, as in case of the time-delay embedding (Section 4.2.2). Since this would result in many combinations, these values are only sampled for odd numbers, i.e. for 1, 3 and 5. At the same time, it cannot be ruled out that in combination with the spatial-neighbor embedding, time-delay embedding will now improve the skill in detecting TCs. Therefore, also the time-delay embedding is varied to the same extent as the spatial-neighbor embedding, i.e. κ is increased to a maximum of 2 and τ is sampled with the values 1, 3 and 5. Altogether the following combinations are analyzed: $\{\kappa_{xy} \in \mathbb{N} \,|1 \leq \kappa_{xy} \leq 2\} \times \{\tau_{xy} \in \mathbb{O} \,|\, 1 \leq \tau_{xy} \leq 6\} \times \{\kappa \in \mathbb{N} \,|\, 1 \leq \kappa \leq 2\} \times \{\tau \in \mathbb{O} \,|\, 1 \leq \tau \leq 6\}$, where $\mathbb{O}$ denotes the set of odd natural numbers.

However, some of these combinations have already been tested in Section 4.2.2 or, as described there, a change of τ or τ_{xy} does not represent another combination if $\kappa = 1$ or $\kappa_{xy} = 1$. As result, there are only 12 new combinations, which have to be tested. The results for these combinations are shown in Figure 4.4. Due to the prime detection skill of $(\kappa, \tau, \kappa_{xy}, \tau_{xy}) = (1, 1, 2, 3)$, the combination $(1, 1, 2, 2)$ was tested additionally, even though τ_{xy} is not an odd number.

In Figure 4.4, one recognizes that even in combination with spatial-neighbor embedding, an increase in time-delay embedding parameters does not improve the detection skill. Again, the skill in detecting TCs gets worse in most cases when increasing κ. There are no synergy effects between spatial-neighbor and time-delay embedding. A change of

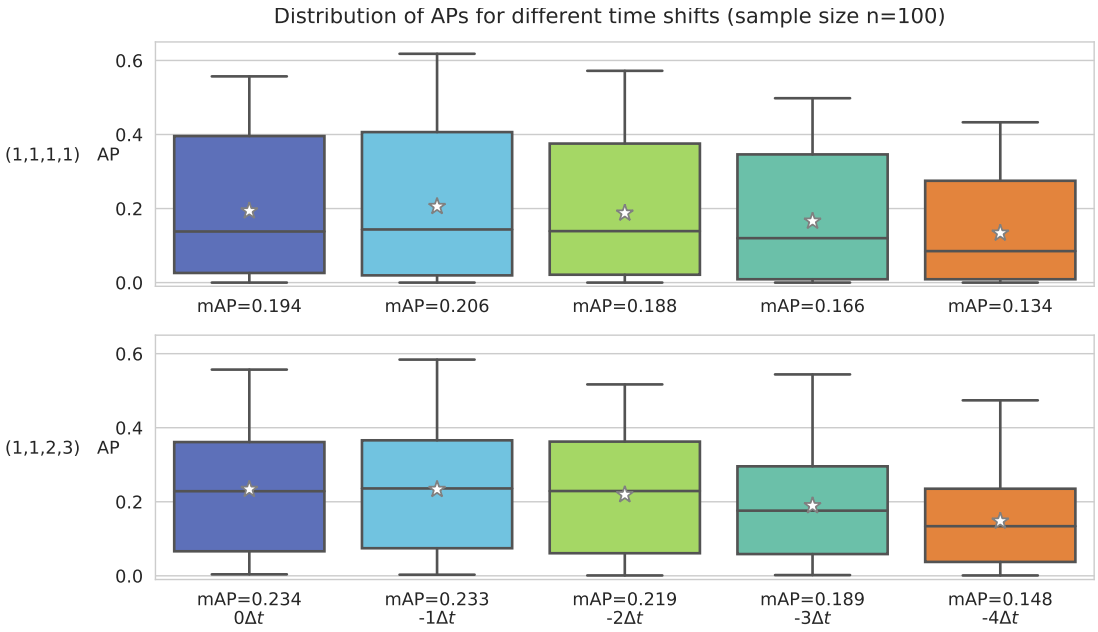


FIGURE 4.5: Box plots showing the distribution of APs in the case of different time shifts Δt. In the upper row results are presented using the embedding parameters (1,1,1,1), in the lower row using (1,1,2,3). Going from left to right, Δt becomes more negative, i.e. the detections are shifted forward by an increasing number of time steps. Selection of the embedding parameters shown is explained in Section 4.2.3.

the spatial-neighbor parameters (i.e. in the figure in the horizontal direction) yields a more differentiated picture: here, the combination (1,1,2,3) performs best with regard to the mAP and is 0.04 higher than the combination (1,1,1,1), i.e. the run without any embedding. When compared in pairs by applying the Nemenyi test, the combination (1,1,2,3) differs statistically highly significant from all other combinations ($p \leq 0.001$), except for (1,1,2,2) and (2,1,2,3) (see Appendix, Table A.1). The maximum AP, the upper end of the whisker, is as high as in (1,1,1,1) and (1,1,2,5). Based on the results, the combination (1,1,2,3) was selected as the preferred one. Due to its lower computational effort and comparable results, especially regarding the maximum AP achieved (the end of the upper whisker), the variant (1,1,1,1) is still pursued. All the following tests are performed concurrently with the two chosen embedding parameter settings.

4.2.4 Extending the detection algorithm by a time shift

Analysis of the detections revealed that they are often slightly delayed in time, i.e. they miss the traversing ground truths. In the following, it is therefore tested whether the detection skill improves if the detections are moved forward by 1, 2, 3 or 4 time steps Δt forward in order to cover the ground truths more accurately. One time step corresponds to 6 h in the ERA-Interim data. This test was performed both with the embedding parameters (1,1,1,1) and (1,1,2,3) concurrently.

The results are shown in Figure 4.5. A shift of the detections affects the skill in detecting TCs. Although the best value of mAP = 0.234 occurs for (1,1,2,3) without time shift, the resulting distribution is not statistically significant distinguishable from (1,1,2,3)

with $\Delta t = -1$ (see Appendix, Table A.1), which achieves mAP = 0.233. Furthermore, in the latter run the upper whisker reaches high values of about 0.6. With larger time shifts the mean values are reduced again, as one can observe in the horizontal direction towards the right in Figure 4.5. A comparison with the row showing the (1,1,1,1) distributions reveals that the mean values here are lower and the distributions become more asymmetrical, i.e. the median and mean value move apart. It is interesting to note that the highest value (the upper end of a whisker) of all runs occurs at (1,1,1,1) and $\Delta t = -1$. Here the upper whisker reaches over 0.6. All in all, it is concluded that a minimal time shift of $\Delta t = -1$ seems to be reasonable, because especially the already well performing variable combinations become even better, which is recognizable by the whiskers becoming longer upwards. A decision between (1,1,1,1,1) and (1,1,2,3) cannot yet be made here, since with (1,1,2,3) $\Delta t = -1$ the mean value and median are higher to such an extent that the Nemenyi test also indicates a significant difference. But the maximum value (the upper end of the whisker) of (1,1,1,1) $\Delta t = -1$ is higher and again the computational effort is lower. Therefore, both parameter combinations are tested when choosing the ideal variable for TC detection (Section 4.3).

4.3 Choosing the ideal variables for TC detection

With the initial settings discussed at the beginning of Section 4.2, as shown in Figure 4.2, the mAP was 0.099 and the maximum AP just under 0.3. With the optimized MDI algorithm settings, mAP is 0.233 and a maximum AP above 0.6. So far, it was not investigated, to which variable or variables the MDI algorithm should be applied to achieve the highest detection skill, but the algorithm was applied to 100 randomly created variable-season combinations. Now the study focuses on choosing the optimal variable(s) by applying the MDI algorithm with optimized parameter settings to all 127 possible combinations. For this, the MDI algorithm is applied to all hurricane seasons from 2000-2010 and the APs obtained are averaged.

As discussed above, the tests are performed with the embedding parameter combinations (1,1,1,1) and (1,1,2,3), which are referred to in the following as setting (a) and (b), respectively.

Figure 4.6 shows the best combinations of variables for each of the two initial settings. Overall, the best detection skill is achieved when using the MDI algorithm univariately on vo850 with setting (a). Here the mAP is 0.537 for the period 2000-2010, but the AP range between individual hurricane seasons is wide, from 0.3 to 0.675. The latter is the best score of all, which is achieved in a single season with a certain combination of variables. The detection skill is worse with setting (b) when looking at vo850, where this combination achieves a mAP of 0.51. It is noticeable that the range of the results for setting (a) is larger for the other variable combinations, i.e. the AP varies more strongly between the seasons. There are also more lower outliers (black diamonds in Figure 4.6) that are no longer within the 3·IQR range. In addition to vo850 and q700, also the variables vo250, z1000 and wind300 appear in the five best performing variable combinations.

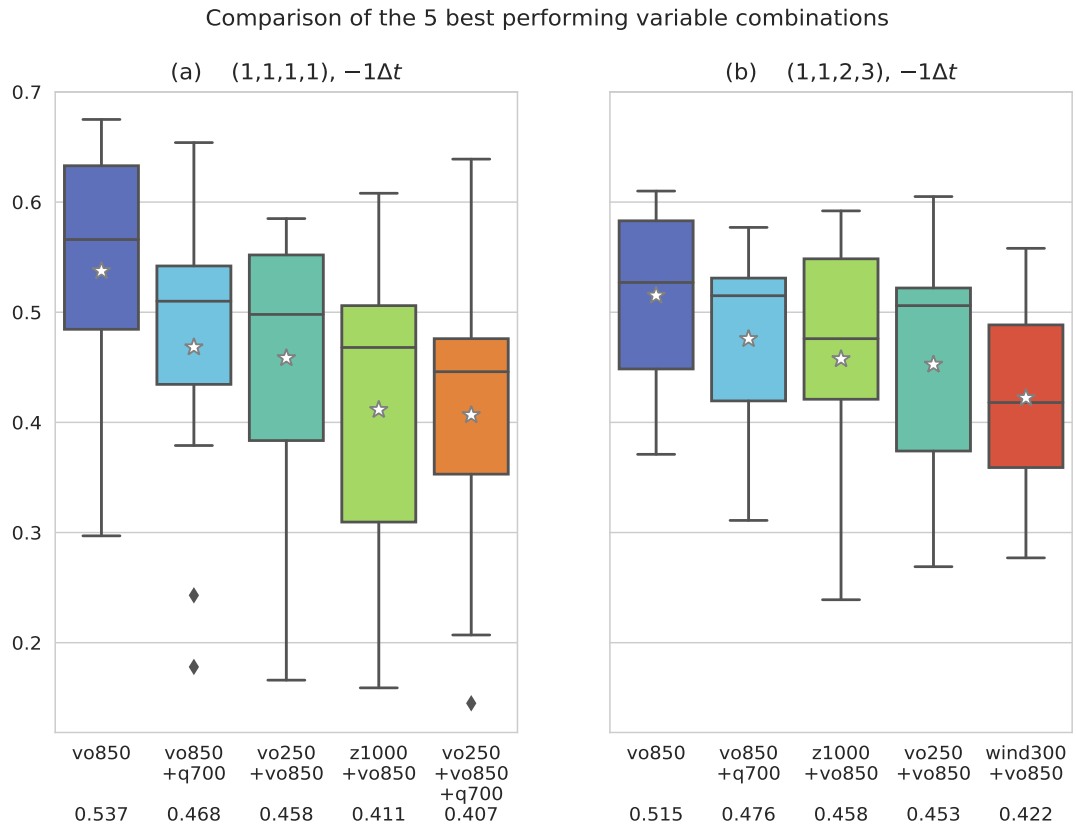


FIGURE 4.6: The five best performing variable combinations for parameter combinations (1,1,1,1) and (1,1,2,3), each with a time shift of $-1\Delta t$. The box plots show the distribution of APs for the hurricane seasons 2000-2010 per variable combination. Below the labels of the x-axis the mAP is given, i.e. the average over all seasons. The mean values are also indicated as stars inside the boxes. If variable combinations occur twice, the corresponding boxes are colored similar.

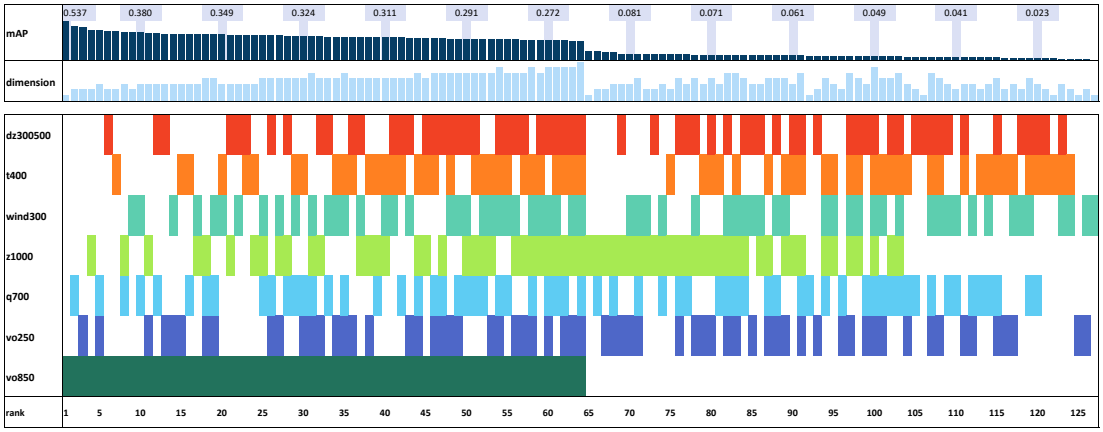


FIGURE 4.7: All 127 variable combinations ranked according to their mAP, which is obtained by the detection of TCs. No embedding was applied, so the embedding parameters (1,1,1,1) were used and a time shift of $\Delta t = -1$ was chosen. On the left one finds the best performing combination, indicated by rank 1, on the right the worst performing combination. The color code above the ranks indicates to which variables the MDI algorithm was applied. In addition, the dimension is indicated by bars, i.e. whether the algorithm was applied univariately (dimension=1) or multivariately to up to seven variables (dimension=7). Above this, the mAP of the combination, i.e. the mean value of the APs from all hurricane seasons, is additionally indicated by bars.

In further plots, the inclusion of individual variables in the ranking of variable combinations is examined. In Figure 4.7, a color code indicates which variables are included in combinations that achieve high ranks and which are further down in the ranking. The ranking is shown for setting (a), but a similar picture is obtained for setting (b) (see Appendix, Figure A.1). It is noteworthy that vo850 is included in all combinations in the top ranks. In the first half of the rankings there is no combination without the inclusion of vo850. The mAP also decreases considerably by 0.277 when vo850 is no longer included in the variable combinations, which can be seen in Figure 4.8. Furthermore, the dimension, i.e. the number of variables to which the MDI algorithm was applied simultaneously, shows a trend and increases from initially 1 (univariate) to the highest dimension of 7, if vo850 is part of the combinations. This can be noticed in the second upper row of the Figure 4.7. Without the inclusion of vo850, starting from rank 65, there is no clear trend anymore, but the dimensions are randomly arranged.
The remaining variables show a more differentiated picture: the inclusion of vo250 is relatively homogeneously scattered, with no pronounced accumulations in the lower or upper ranks. Including this variable reduces the detection skill by a mAP of 0.003 slightly, as shown in Figure 4.8. q700 is included in some of the upper ranks and there is no inclusion at least in the lower ranks in Figure 4.7. The inclusion of q700 improves detection skill by a mAP of 0.008 (Figure 4.8). With z1000 one observes that it has a leading role especially in the part of the ranking without vo850, i.e. in the lower half of Figure 4.7. Approximately in the lowest fifth of the ranks no combination containing z1000 can be found. On average, the inclusion of z1000 improves detection skill by a mAP of 0.031 (Figure 4.8). The inclusion of wind300 is again rather randomly scattered in Figure 4.7 but is the last variable that is involved seen from the first ranks. All other variables appear earlier in combination with other variables or univariately. The variable is not beneficial to overall detection skill and reduces the mAP by 0.019 on average (Figure 4.8). T400 and dz300500 tend to rank lower, especially when the two halves of Figure 4.7 with and without vo850 are considered in isolation. They are hardly represented in the upper ranks. On average the inclusion of T400 lowers the mAP by 0.008 (Figure 4.8), the inclusion of dz300500 by 0.005.

As a conclusion of this section and in particular Figure 4.6, it is concluded that a univariate application of the MDI algorithm on vo850 with the setting (a) (1,1,1,1) $-1\Delta t$ is best suited to detect TCs.

4.3.1 Relative vorticity on alternative pressure levels

Due to the promising results when applying the MDI algorithm to vo850, the algorithm is tested on the relative vorticity at additional pressure levels. The embedding and time shift settings are the same as before, i.e. (1,1,1,1) and $-1\Delta t$. The MDI algorithm is again applied to the hurricane seasons 2000-2010. The goal is to investigate if the detection skill can be further improved by using other pressure levels than 850 hPa for relative vorticity. Near the 850 hPa pressure level, particularly close-meshed tests were

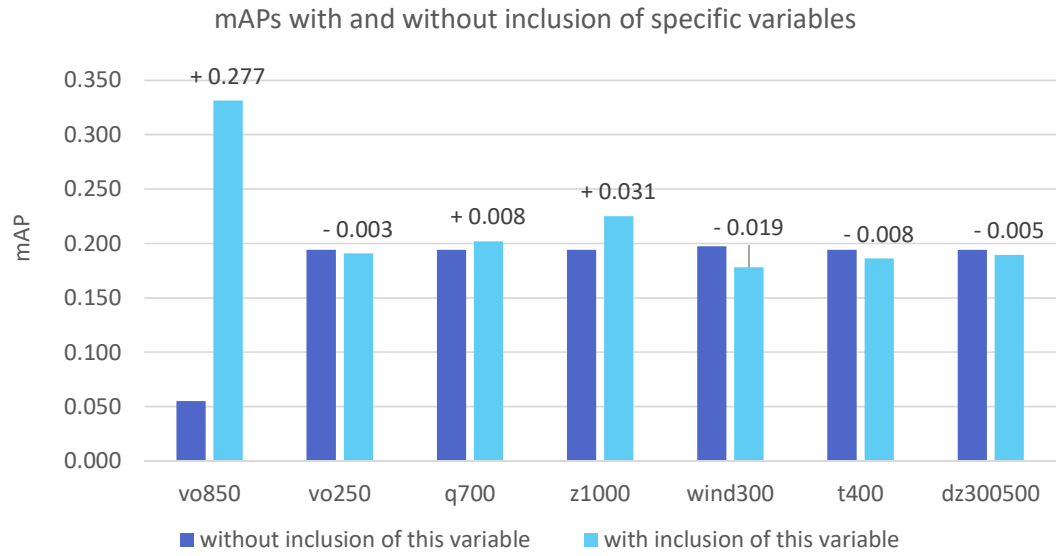


FIGURE 4.8: Averaged mAP when using setting (a) without and with the inclusion of a specific variable. The values are calculated by averaging the mAPs of the 63 runs without this variable respectively the 64 runs with this variable. The numbers above the light blue bars show the difference in mAP without and with inclusion of the variable.

carried out in intervals of 25 hPa. The distributions are shown in Figure 4.9 using box plots. As one can see, vo850 still achieves the best detection skill: the mean value is the highest compared to the other pressure levels and the whiskers cover a similar range as for the other pressure levels. Using relative vorticity at decreasing pressure levels (in greater atmospheric height) as input variable, decreases the detection skill. When using relative vorticity at 200 hPa, the MDI shows the worst detection skill. When applying the MDI algorithm to the relative vorticity at pressure levels adjacent to 850 hPa, the detection algorithm shows a similar skill, but the averages are slightly lower than at 850 hPa. When applied to vo700, the mAP is 0.489 which is also close to the mAP of 850 hPa. Additionally, the spread of the distribution of vo700 between the individual hurricane seasons is small (0.25), as can be seen from the short whiskers. No other level with a mAP above 0.02 has a similar small spread. The 2007 hurricane season, which causes problems when applying the MDI algorithm to the relative vorticity and mostly represents the most extreme outlier downwards (see the blue bars of Figure 4.10), is less pronounced here: The AP here is about 0.38.

It can be concluded that 850 hPa is the pressure level at which the MDI algorithm shows the best performance in TC detection when applied to relative vorticity.

4.4 Improvement of detection skill through post-processing

The MDI algorithm returns a list of the detections ranked by their score determined using the U-KL method. Post-processing allows to further improve the detection skill. There are two ways to achieve this:

- Filtering the results with soft thresholds, mainly to sort out false alarms on the

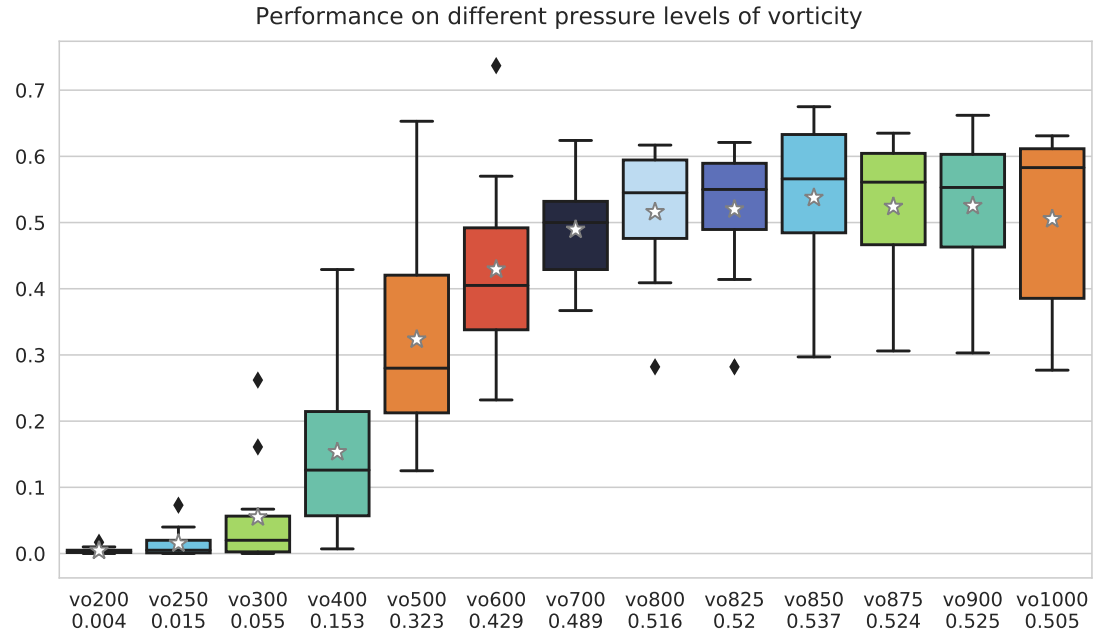


FIGURE 4.9: Results of the application of the MDI algorithm to the relative vorticity at additional pressure levels. The respective mAP, averaged over all hurricane seasons analyzed, is given below the x-axis.

upper ranks of the returned list (see Section 3.2) and thus improve the AP.

- Applying a cutoff threshold: assuming that all detections with a score above a certain threshold are TCs, it can be defined up to which rank the list of detections is read. All detections with a lower score than the threshold are discarded.

First, the results by applying soft thresholds are analyzed.

4.4.1 Soft thresholds

The three filters introduced in Section 3.5 are now applied to the list of detections. The thresholds and parameters of the filters are varied as summarized in Table 4.2.
It applies to all filters that the given conditions (see first column of Table 4.2) must be valid during at least half of the time steps of a detection. The possible filter configurations result in $9 \cdot 5 \cdot 6 = 270$ different combinations, which can be applied to the list of detections obtained by applying the MDI on vo850 in the hurricane seasons 2000-2010. Table 4.3 shows the 5 best combinations of filter settings and their resulting mAPs as well as the rank and the result without filtering and with the filtering that leads to the worst mAP. With a mAP of 0.555, the combination with a wind speed threshold of $8\,\mathrm{ms}^{-1}$, a relative vorticity filtering of $3 \cdot 10^{-5}\,\mathrm{s}^{-1}$ and no pressure filtering achieves the best result. For comparison, the mAP without filtering is 0.537. However, the thresholds in wind speed and relative vorticity are high with regard to the traditional thresholds (see Section 2.3). Furthermore, overtuning on the training data set should be avoided, i.e. the thresholds should not be chosen in such a specific way that they are only applicable to a certain region, resolution and time period.

TABLE 4.2: Overview of the filters and the range of parameters tested to determine the optimal soft thresholds.

Filter	Sample range	Unit or meaning
Wind filter: checks whether a given wind speed at the surface is exceeded at least once in the detection at a time step.	0, 2, 4, 6, 8, 10, 12, 14, 16	ms^{-1}
Relative vorticity filter: checks whether the required relative vorticity at 850 hPa is exceeded at least once in the detection at a time step. Furthermore, the relative vorticity averaged over all grid points at one time step must be positive.	0, 1, 2, 3, 4	$10^{-5}\,\mathrm{s}^{-1}$
Pressure filter: checks whether the average geopotential height at 1000 hPa within the detection box is lower than that within a surrounding box; the extent of the surrounding box in number of grid cells is used as the filter parameter.	0, 1, 2, 3, 4, 5	Grid spacings that the surrounding box is larger on each side

TABLE 4.3: The 5 best combinations of filter settings, sorted by their mAP calculated by averaging over 11 hurricane seasons. Furthermore, the rank and the mAP without filters, as well as the filtering with the worst result on rank 270 are displayed.

rank	wind speed threshold in ms^{-1}	relative vorticity threshold in $10^{-5}\,\mathrm{s}^{-1}$	pressure filter setting in number of grid cells	mAP
1	8	3	0	0.555
2	8	2	0	0.555
3	8	1	0	0.555
4	8	4	0	0.550
5	2	4	2	0.545
⋮	⋮	⋮	⋮	⋮
80	0	0	0	0.537
⋮	⋮	⋮	⋮	⋮
270	16	4	3	0.285

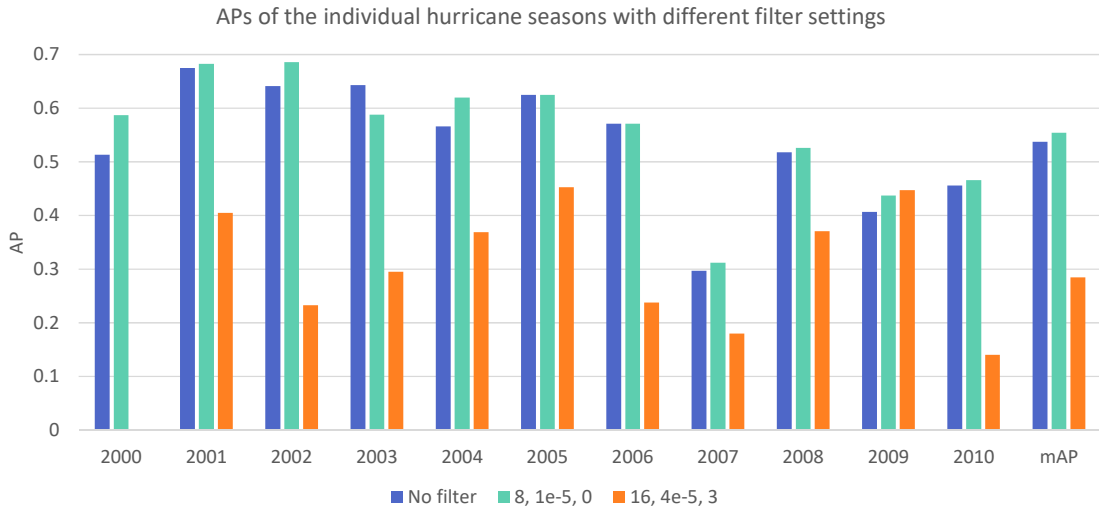


FIGURE 4.10: Bar plot illustrating the effect of filters on the AP for each season. The default configuration without filters is shown in dark blue, the optimal filter setting in turquoise and the worst performing configuration in orange. The bars on the very right show the average over all years. In the legend, the first number represents the wind filter, the second the relative vorticity filter and the last one the pressure filter.

The mAPs in the upper five ranks of Table 4.3 change only by 0.01. It is noticeable that the wind speed thresholds fluctuate, but never exceeds $8\,\mathrm{ms}^{-1}$. Moreover, the relative vorticity threshold varies, and even reaches the upper end of the tested range with $4 \cdot 10^{-5}\,\mathrm{s}^{-1}$. The comparison with the surrounding geopotential height is not a suitable filter, as its parameter is 0 in most of the upper ranks, i.e. the filter is turned off. It is concluded from these results that filtering using thresholds for wind speed and relative vorticity can be useful. It is not crucial which exact thresholds in certain ranges are selected for these two parameters. Regarding the applicability of the algorithm to other resolutions and other regions, rather relaxed values with a relative vorticity threshold of $1 \cdot 10^{-5}\,\mathrm{s}^{-1}$ and a wind speed threshold of $8\,\mathrm{ms}^{-1}$ are chosen. The latter value is still conservative compared to existing methods. Walsh et al. (2007) recommend a wind speed limit of about $16\,\mathrm{ms}^{-1}$ at a horizontal resolution of about 75 km, which is the one of ERA-interim data.

On basis of Figure 4.10, it is examined how the filtering influences the results for the individual seasons. In most seasons, the optimal filtering (wind speed threshold of $8\,\mathrm{ms}^{-1}$ and relative vorticity threshold of $1 \cdot 10^{-5}\,\mathrm{s}^{-1}$) raises the AP, except for the seasons 2003 and 2006. The extreme filtering (wind speed threshold of $16\,\mathrm{ms}^{-1}$ and relative vorticity threshold of $4 \cdot 10^{-5}\,\mathrm{s}^{-1}$) shown in orange, which leads to the lowest mAP, reduces the AP significantly in most cases, even to 0 in the season 2000. However, in a single case, the season 2009, this filter setting increases the AP.

In order to better understand how AP is affected by filtering, the exemplary precision-recall curves of the years 2000, 2003 and 2005 are considered before and after the best possible filtering (Figure 4.11). In these cases, the filtering has an improving, a reducing and hardly any effect on the AP of the respective season. When comparing the light and dark blue curves of the 2000 season, it is noticeable that some false alarms were sorted out from the upper ranks, which shifts the filtered curve to the right compared to

TABLE 4.4: Score thresholds calculated to ensure that the detector has a precision of at least 0.8 in each season.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Threshold	1156	2238	899	678	480	370	1958	6249	574	5286	1237

the unfiltered curve, so that a larger area is created under the curve. This is equivalent to an increased AP. Lower down in the ranking, however, there is still a true detection in both lists, which can be seen by the upward jag of the AP curve at a POD of approx. 0.6. The light and dark green curves represent the 2003 season before and after filtering. Here, the application of the filters leads to a decrease in AP, since some true detections are filtered out. Consequently, the dark green curve is shifted to the left compared to the light green curve. In contrast, there is only little effect of the filtering in the 2005 season as can be seen from the very similar orange and red curves.

4.4.2 Defining of a cutoff threshold

As one can see from the curves in Figure 4.11, the intersection points between the AP curves and the precision line at 0.8 move closer together when filtering is applied. By this, a more uniform POD with a mean precision of 0.8 over all the hurricane seasons is obtained.
The exact value of the precision line could be chosen differently, but 0.8 seems to be a compromise between precision and POD. Many hits are kept, which can be seen in the mostly horizontal AP curves in Figure 4.11 above a precision of 0.8. Below a precision of 0.8 there are many false alarms and little hits. This can be seen from some "saw teeth" of the curves, but an overall vertical downward path of the curves.
Using the intersection points of the AP curves and the 0.8 precision line, one can now define a score threshold above which the detections are kept and below which they are discarded. This is possible by selecting the value pair of PRC and POD with the highest POD each season, which is still above the 0.8 precision line. This value pair represents a detection in the list returned by the MDI algorithm. This detection received a score from the MDI algorithm. From these resulting scores in each season, the median is chosen to receive a final score below which the detections can be discarded.
As can be seen in Table 4.4, the score threshold varies greatly between 370 and 6249. As a simple approach, the median is chosen because the value should to be as robust as possible against seasons in which the cutoff threshold is very high, such as in 2007. This results in a threshold of around 1150 below which the detections will be discarded.

4.4.3 Final detection skill and detection examples

In this section, the detection skill of the MDI algorithm with optimal configurations and post-processing will be examined. The detection skill is presented in Table 4.5. When searching for TCs with wind speeds $\geq 34\,\text{kn}$, i.e. tropical storms, the detector reaches a mAP of 0.452, which is slightly below the mean POD of 0.455. This corresponds to about half of the ground truths being detected. The false alarm rate is 0.162, which

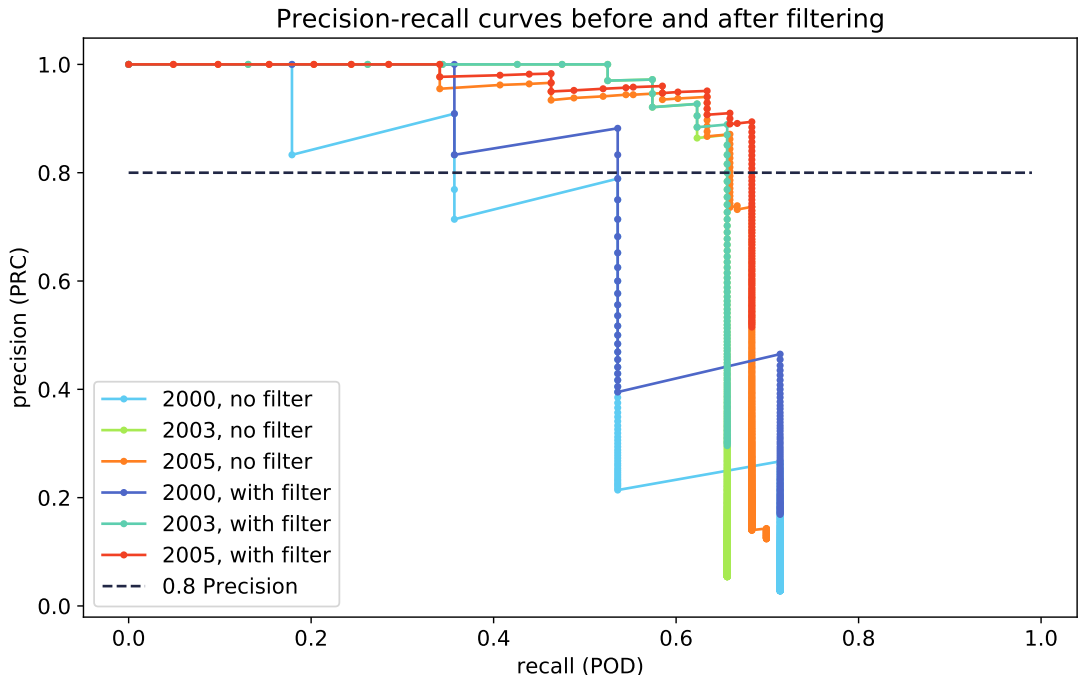


FIGURE 4.11: Precision-recall curves of selected seasons before and after filtering. The optimal configuration from Section 4.4.1 with a wind speed of $8\,\mathrm{ms}^{-1}$ and a relative vorticity of $1 \cdot 10^{-5}\,\mathrm{s}^{-1}$ is applied. The dashed line shows a precision of 0.8.

means that about one out of 6 detections does not cover a ground truth at all or sufficiently. The average precision is 0.838. The BIAS has a value of 0.539, but is not meaningful in this context, as will be discussed in Chapter 5.5. One reason for this is that single detected ground truths do not indicate that a storm has been completely detected, since the storm track is composed of several ground truths. 272 ground truths were detected, 316 cases were missed out. 27 detection boxes are false alarms.

TABLE 4.5: Performance of the detector in its final configuration when applied to the Gulf of Mexico region for the hurricane seasons 2000-2010. The values mAP, POD, FAR, PRC and BIAS represent averages over all seasons.

mAP	POD	FAR	PRC	BIAS	Hits	False alarms	Misses
0.452	0.455	0.162	0.838	0.539	272	27	316

Figure 4.13 shows some examples when applying the detector to relative vorticity at 850 hPa. Particularly strong storms are easily detected by the MDI algorithm, such as hurricane Katrina in 2005. Severe storms are usually covered by a detection box that reaches the predefined maximum size of 8 grid spacings. Less severe storms are often enclosed by non-square and small detection boxes. This is also the case if the detections are located at the edge of the investigated region, as it is the case with storm Stan in 2005. The detections of particularly strong storms such as the hurricanes Rita and Katrina are listed high in the ranking, as can be seen from the low, small numbers inside the detection box (e.g. 1 for hurricane Katrina) in Figure 4.13. Less intense storms such as Stan and Charley are further down in the ranking and are marked by higher numbers.

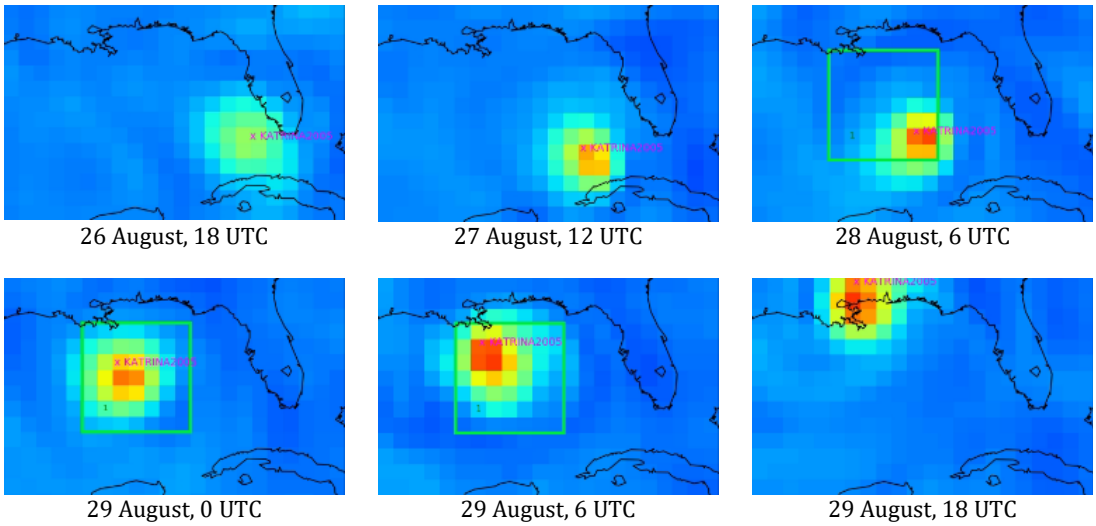


FIGURE 4.12: Detection of hurricane Katrina in 2005. The description of the image corresponds to Figure 4.13, except that a temporal sequence of a detection is shown here.

Looking at the false alarms in Figure 4.13, one can see that these cases are mostly storm-like events with increased relative vorticity values that do not reach the strength of a TS. Examples include tropical depressions or unnamed events. As in the case of hurricane Rita in 2005, sometimes peripheral areas of storms are detected. Since the ground truth usually represents the storm center, detections of peripheral areas are not considered as hits. Sometimes a ground truth passes through the detection box too fast as in the case of Dean in 2007. In the evaluation, detections are only considered as hits if the detection box contains a ground truth for at least 25 % of the time steps. In some cases, the classification as false alarm is therefore not completely justified and is rather a result of the quite strict definition of a hit, since the MDI algorithm correctly detects areas of increased relative vorticity.

In the examples shown, all misses are weak storms, often near the edge of the investigated region. Thus, they are only present for a short period of time and the MDI algorithm does not detect them or rank them low. As a result, they are sorted out by the score thresholds in the post-processing step.

Figure 4.12 shows a sequence of images covering part of hurricane Katrina's life cycle over the Gulf of Mexico and the corresponding detection box. As can be seen, the storm is detected by the MDI algorithm only once it reaches a certain strength. This is the case on 28 August, 2005. The storm then passes through the detection box for a few time steps before leaving it. After this, the detection box disappears. On the evening of 29 August, the storm center is located entirely over land. Here hurricane Katrina is no longer enclosed by a box. Consequently, the storm is only detected by a single box that is statically placed while the storm is passing through.

The MDI algorithm places the storm high in the ranking of the season 2005, which can be seen from the small 1 inside the detection box. Only the even stronger hurricane Rita in the same year receives a higher ranking, which can be seen from the 0 in the detection box of Figure 4.13.

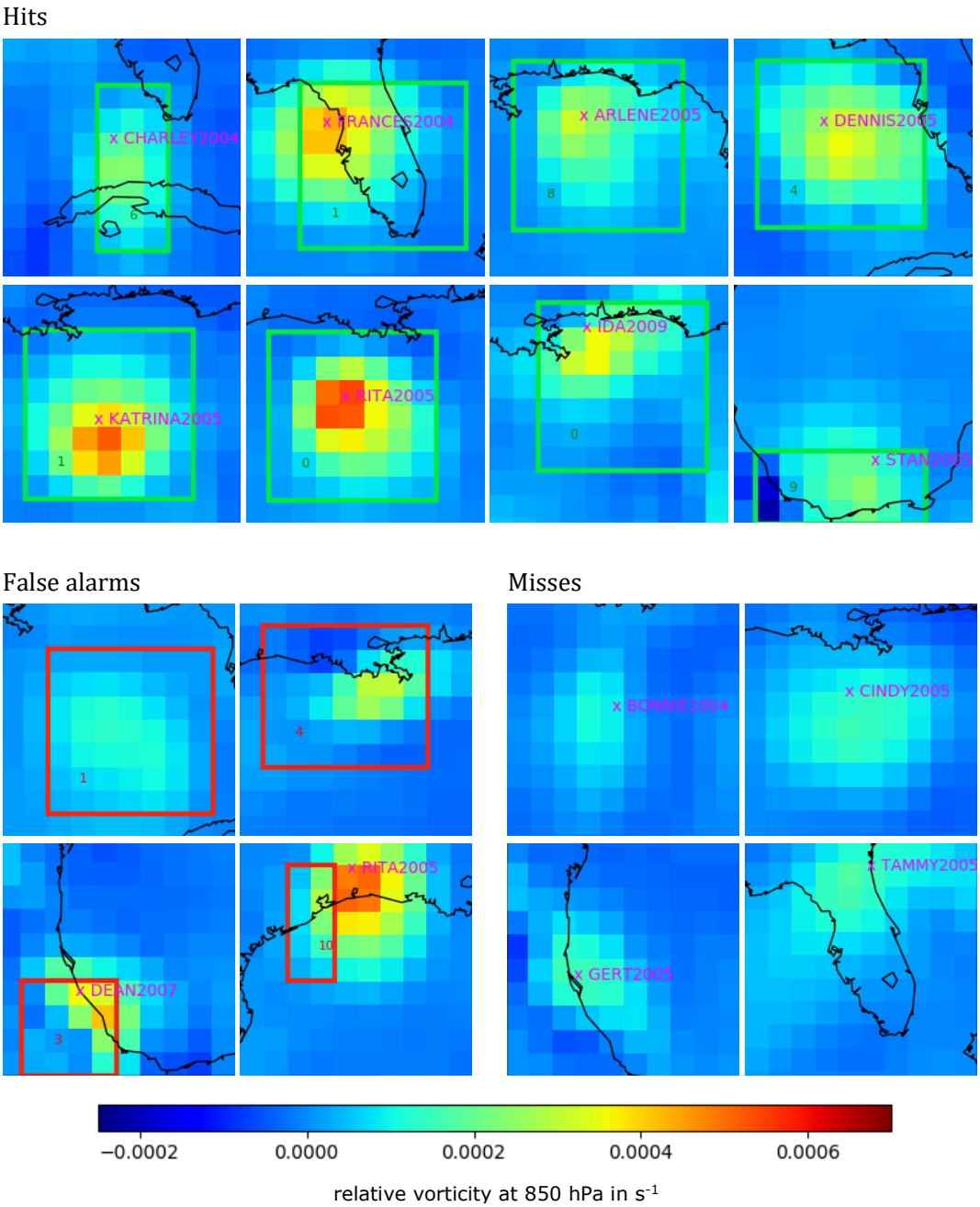


FIGURE 4.13: Examples of hits, false alarms and misses using the final detector configuration when applied to the relative vorticity at 850 hPa. Ground truths are all entries of the IBTrACS database with a wind speed $\geq$ 34 kn, i.e. when the event reaches the strength of a tropical storm (TS). In addition to the data in the background and the coastlines, detection boxes are visible: green boxes show detections that have been rated as hits, red boxes represent false alarms. The ground truths are marked by a pink X together with the storm name and year of occurrence. The small number inside the detection box indicates the rank on which the detection is placed by the MDI algorithm in the returned list based on its score.

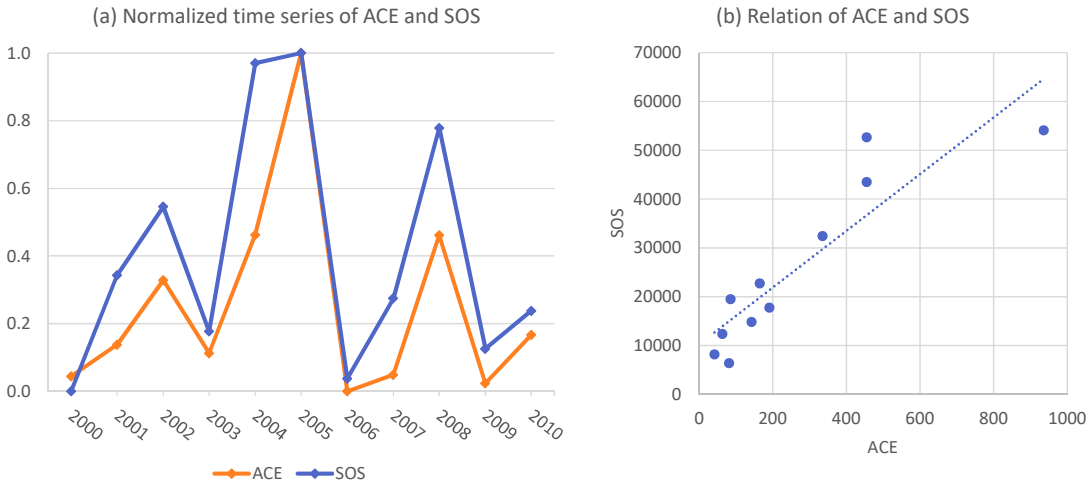


FIGURE 4.14: Correlation of ACE and SOS: (a) shows the normalized ACEs and SOSs of the individual seasons along a time axis. In (b), ACE and SOS of each season are plotted as a scatter diagram. In addition, the linear regression curve is drawn.

4.5 Correlation of strength of a hurricane season with scores

In addition to the direct detection of individual TCs, it is also possible to draw further conclusions from the application of the MDI algorithm: scores calculated by the MDI algorithm can be used to estimate the strength of a given hurricane season by adding up all score of that particular season. This measure is called *sum of scores* (SOS) in the following. Another measure to assess the strength of a TC or an entire hurricane season is the accumulated cyclone energy (ACE) (see Section 2.1.1). In the case of a strong correlation between these two quantities, the MDI algorithm could estimate the strength of individual hurricane seasons. By applying it to future projections from climate models, potential changes of the TC activity in a changing climate could be assessed.

In this thesis, ACE is calculated based on all ground truths in the IBTrACS database with a wind speed $\geq 34\,\text{kn}$, which are present in the investigated area of the Gulf of Mexico.

Figure 4.14 shows the results of the comparison between ACE and SOS. Visually, the progression in (a) is comparable for both curves, even if there are some deviations in individual seasons. The biggest difference can be seen in the 2004 season, which, according to ACE, has only a medium strength, whereas SOS classifies the season as particularly active. For the seasons 2000, 2003, 2005 and 2006, SOS and ACE show consistent seasonal activities. Pearson's correlation coefficient for the two quantities is $r = 0.9$ (see Figure 4.14 (b)). With a p-value of $1.65 \cdot 10^{-4}$, which was determined by double-sided testing of the t-distribution, this value is also statistically significant. The regression line is approximately given by

$$\text{SOS} = 58 \cdot \text{ACE} + 10280. \tag{4.1}$$

The equation shows that even in the absence of any storm activity (ACE = 0), the MDI

algorithm returns detections with a sum of approximately SOS = 10280.

5 | Discussion

In the previous Chapter 4, the optimal initial settings of the MDI algorithm were determined, detection skill was improved by post-processing measures, and the detector was evaluated. In this chapter, the individual measures in preprocessing (Sections 5.1 and 5.2) as well as post-processing (Section 5.4) are evaluated. Furthermore, the choice of the variable combination is discussed (Section 5.3) and the final detection skill is assessed (Section 5.5). The individual tests will be analyzed in the order in which they were presented in Chapter 4.

5.1 Variables and initial settings

In an initial case study, variables at different pressure levels were first selected for univariate and multivariate application of the MDI algorithm. This was done to limit the number of possible variable combinations, but there may be other variables that could be used to detect TCs. For example, the OWZ parameter (Tory et al. 2013b) might be considered, which measures large-scale, favorable conditions for formation of TCs and is therefore less dependent on model resolution compared to other variables or criteria, such as the presence of a warm core.
By using a different modelling method for the variable distributions, such as KDE (see Section 3.3.2), parameters like total precipitable water, precipitation or cloud cover that are not normally distributed could also be investigated. Furthermore, the restriction of variables to certain pressure levels could be avoided by applying the MDI algorithm spatially not only two-dimensionally, but three-dimensionally to the whole troposphere. This would increase the computational effort, but it could lead to better detection skills. This is because anomalies that are limited only to a shallow layer of the troposphere would receive a lower ranking than TCs that span a large part of the troposphere vertically. This would be similar to the idea of using criteria that average or combine variables of multiple pressure levels. As shown in Section 2.3, this approach is sometimes used in TC detection (e.g. Bengtsson et al. 1995, 1996).

The finding that the U-KL divergence turns out to be the best of the investigated measures to detect TCs underlines the results of Barz et al. (2018), who used the MDI algorithm together with this divergence method for the detection of low-pressure systems in spatio-temporal data. Still, this was somehow unexpected, since in another example by Barz et al. (2018) the cross entropy is better suited to detect fast moving objects in videos and the U-KL shows better detection skill in detecting particularly calm

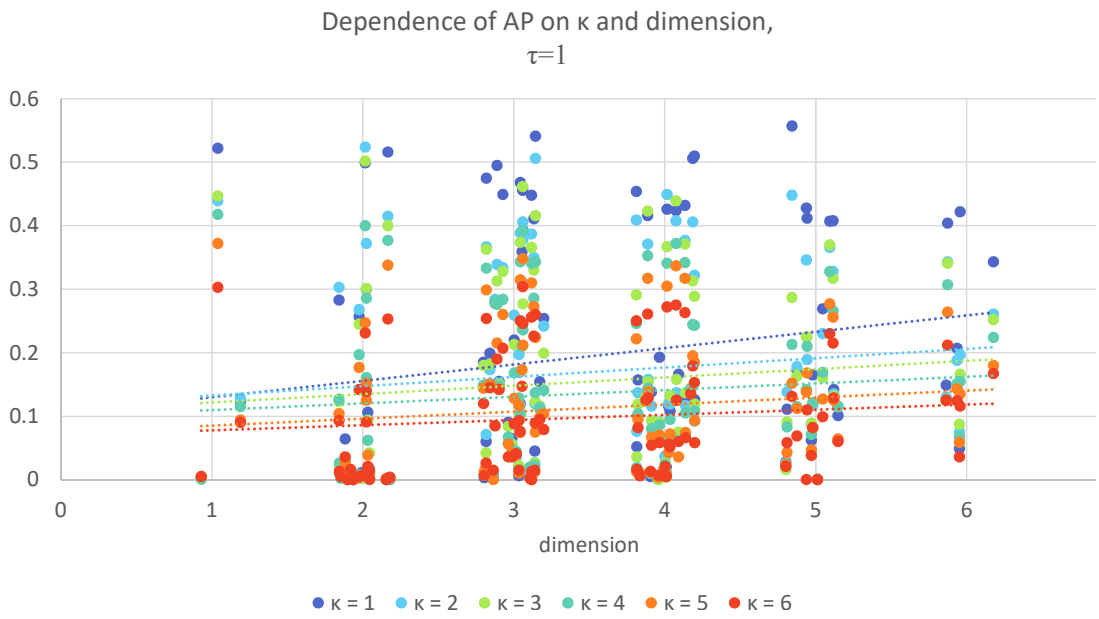


FIGURE 5.1: Dependence of AP on κ and dimension. Detailed representation of the first line of Figure 4.3. The individual points represent the respective AP of the various variable combinations, which are grouped by dimensions (1 to 6). The different colors represent the results when using different values of κ.

intervals. The poor detection skill of the KL divergence (i.e. the biased version of the U-KL divergence) is explained by many small, incoherent detections. Here, anomalies other than TCs move up in the ranking of the MDI algorithm and thus reduce the AP. The algorithm does not recognize the exceptionally large, coherent TC anomaly, but splits such events into many small detections and, at the same time, dilutes the ranking with other spatial and temporal small-scale, yet strong anomalies.

5.2 Embedding parameters

The results of the search for the optimal time-embedding parameters are also partly unexpected. Initially, it was expected that incorporation of surrounding data points would improve detection skill and that medium values for parameters such as $\kappa = 3$ and $\tau = 2$ would yield the best results. Barz et al. (2017) also used moderate time-delay embedding parameters for similar applications ($\kappa = 3$, $\tau = 1$). However, the best result is achieved for $\kappa = 1$ and $\tau = 1$, i.e. without any time-delay embedding. The first hypothesis to explain this finding is that there might be enough data available in the analyzed intervals and that it is not necessary to extend the data by incorporating further surrounding data. By including additional data points, other long-term and rather hidden anomalies may be detected more often than TCs.

When looking at the first row of Figure 4.3, it is noticeable that the median only changes by 0.02 and the mean value changes by 0.05. This could indicate that the AP of well performing combinations is reduced most by increasing κ values.

From Figure 5.1 it can be seen, that the AP of higher-dimensional variable combinations is reduced more with an increasing κ than the AP of low-dimensional ones. This

can also be seen by the slope of the regression lines decreasing with increasing κ. This means, under an increasing κ, the AP is reduced more at higher dimensional variable combinations than at lower ones. The reason could be the following mechanism: if, for example, 6 variables are already involved, the time series for $\kappa = 3$ is 18-dimensional and consequently the amount of data points for estimating such a high-dimensional distribution is no longer sufficient.
Another explanation for the poor detection skill with higher time-delay embedding parameters could be the time resolution of the ERA-Interim data, which is 6 h. During this time, TCs sometimes move over large distances, which is why incorporating data from previous time steps tends to wash out the anomaly rather than to provide any benefit.
The range of tested κ and τ combinations can be considered sufficient, since Figure 4.3 shows that each increase in these parameters leads to decreased detection skills.

The results of the search for the optimal spatial-neighbor embedding parameters underline the findings that no time-delay embedding should be applied. Even if the spatial-neighbor and time-delay embedding parameters are changed simultaneously, detection skill is reduced when the latter are increased.
However, a certain spatial-neighbor embedding with $(\kappa, \tau, \kappa_{xy}, \tau_{xy}) = (1, 1, 2, 3)$ as parameters has positive effects, especially on poorly performing variable combinations, since the median of the APs increases compared to the tests without this embedding, as can be seen in Figure 3.4. There seems to be little change in the already well performing combinations, as the upper whisker has the same upward extension as for the test without embedding. In this experiment, the combination (1,1,2,4) could have been tested for completion, since it is also similar to the best performing combination (1, 1, 2, 3).

5.3 Commonly used and new variables

This study also examined to which variable(s) the MDI algorithm should be applied in order to detect TCs most reliably. Here the relative vorticity at 850 hPa stands out. This confirms many traditional methods (e.g. Bengtsson et al. 1982; Haarsma et al. 1993; Vitart et al. 1997; Walsh and Watterson 1997), which also work with this variable on the same pressure level. Contrary to initial expectations, detection skill is reduced when other variables are added. The reason for this is difficult to pin down. When using the MDI algorithm, the increased relative vorticity seems to be the unique characteristic of TCs compared to other anomalies. This is also underlined by Figure 4.7, which shows that the relative vorticity at 850 hPa is apparently essential when looking for TCs, since any combination of variables scores better when vo850 is included than without.
Moreover, the increased mid-level humidity, represented by q700, seems to indicate TCs, as Briegel and Frank (1997) have already mentioned in general as a basic condition for

TCs formation (see 2.1.2). Geopotential height respective air pressure anomalies probably only indicate TCs to a certain extent in this study – in many cases they are possibly too unspecific. However, in the absence of vo850, geopotential height anomalies, represented by z1000, are the best indication for TCs.
The wind speed at 300 hPa provides the MDI algorithm with no useful information for detecting TCs: if the MDI algorithm is applied to a variable combination with wind300, this lowers the mAP the most (see Figure 4.8).
The rather poor detection skill of variables representative for a warm core (t400 and dz300500) could be explained by the fact that warm cores are not sufficiently represented in the available resolution. However, visual checks of the data showed that the warm core is visible. The magnitude of wind speed and relative vorticity are also reduced by the lower resolution, but the latter is still a prime indicator for TCs using the MDI algorithm. A better explanation is probably the small spatial extent of the warm core, which comprises only a few grid cells. Thus, detections containing the warm core achieve a relatively low score compared to large-scale temperature anomalies that do not represent TCs.
Bases on these findings, vo850 with the embedding parameters (1,1,1,1) has been selected as the preferred variable-parameter combination in this study. This choice is debatable, because vo850 with (1,1,2,3) shows similarly positive results. Additionally, AP fluctuates less over different seasons with the latter combination, i.e. the whiskers in Figure 4.6 are shorter. The disadvantages are lower maximum APs and a more than fourfold increase in computing time: on the preprocessing partition of the computing system MISTRAL of the German Climate Computing Center (Deutsches Klimarechenzentrum (DKRZ)), the total computing time is 393 s (on average about 36 s seconds per season). Without embedding, i.e. with the parameters (1,1,1,1), MISTRAL computed a total of 91 s or about 8 s seconds per season. Especially when applying the MDI algorithm to global data sets, longer time series than in this study or data with high temporal or spatial resolution, this difference in computational effort is relevant. All in all, the choice of the embedding parameters (1,1,1,1) therefore seems justified.

The results shown in Figure 4.9 were also used to investigate whether the MDI algorithm could be applied to other variables in which a TC has an easy to identify signature. Regarding relative vorticity, many of the studies published so far justifiably use thresholds of this variable at 850 hPa. Using directly adjacent pressure levels lead to a very similar, yet slightly worse detection skill. There are few papers that use other pressure levels for relative vorticity than 850 hPa: Wu and Lau (1992) apply the condition that relative vorticity must be positive at 950 hPa for TCs, Tsutsui and Kasahara (1996) use the same condition at 900 hPa. Although 950 hPa was not explicitly investigated in this study, the findings suggest (see Figure 4.9) that this variable is also quite suitable for identifying TCs, since the adjacent levels 900 and 1000 hPa show positive results. No other study known to the author uses further thresholds at other pressure levels. Otherwise, there are only comparing criteria: in Bengtsson et al. (2007) the difference

between the 850 and 250 hPa relative vorticity must exceed a certain threshold and all levels in between must have a positive relative vorticity. Strachan et al. (2013) and Caron et al. (2013, 2011) use the criterion that relative vorticity at 850 hPa must be greater than at 250 hPa. This makes sense because the relative vorticity at 850 hPa shows anomalies in the presence of TCs in this study, characterized by exceptionally high relative vorticity values, whereas the 250 hPa relative vorticity shows hardly any anomalous behavior when influenced by TCs, i.e. will be mostly close to normal.
An interesting approach, as mentioned at the beginning of the discussion, is to apply the MDI algorithm to several relative vorticity levels at once in order to use it spatially in three dimensions and not to consider only one level. This approach could be promising, since already the combination of vo250 and vo850 ranks among the five best performing variable combinations, as shown in Figure 4.6. However, it is questionable whether this combination of many relative vorticity levels adds any new information or, as already shown in Figure 4.7, further variables added to vo850 only reduce the detection skill. The exact effect needs to be investigated in further studies.

5.4 Impact of post-processing

The obtained results when searching for the optimal filtering show one thing above all: mAP is not very sensitive to the exact filtering values selected for wind speed and relative vorticity. It is more important that soft thresholds are used at all. The main purpose of these is to filter out other anomalies, such as calm ones. However, mAP can be increased rather little by these measures. This was also expected, since no new hits are added, but only individual false alarms are sorted out. The pressure filter seems to be generally less useful; it only provides an improvement in AP in certain seasons. This could be related to the fact that it uses surrounding grid points, as shown in Figure 3.6: some detections are located at the edge of the investigated region. Here the surrounding values are missing on one or two sides. In this case, the filter works less reliably than in the case of storms that occur in the middle of the investigated region. Detections in the middle of the investigated area are filtered in a more reliable way, since all four surrounding sides can be taken into account here. Accordingly, seasons in which detections occur preferentially in the center of the investigated region experience a slightly different filtering than seasons in which detections occur mainly at the edge. All in all, filtering of the detections is useful to ensure that anomalous calm intervals are sorted out. This filtering also does not contradict the basic idea of avoiding classical thresholds, as the filters are conservatively chosen.

When selecting a cutoff threshold, there are many alternative approaches to define it. The procedure used in this study – choosing the median of the proposed thresholds from each season – is a first approach. An important basic assumption is that the scores of the detections are independent of the strength of the hurricane season. In other words, a category 5 hurricane in a relatively calm season should receive a similar score as in an active season. This is only approximately the case, since the score, i.e. the U-KL

divergence, is calculated using the remaining data of a season. Therefore, if the season is very active, the score for a category 5 hurricane decreases slightly compared to a calm season. Since TCs are rare events even in active seasons, this effect is not expected to be very strong.
A different approach avoiding this problem is to define a relative cutoff threshold: the distribution of scores in a season is considered and, for example, all detections with a score not in the upper quartile could be discarded.
Moreover, a precision of 0.8 does not necessarily have to be used to derive a cutoff criterion. Another method is to select the score threshold in such a way that the actual number of observed TCs is reproduced as accurately as possible in each season. In the case of 5 storms in the investigated area per season, the MDI algorithm should also return 5 detections. In each season, the detection list of the MDI algorithm must be terminated at a different rank, i.e. at a different score. Subsequently, the median of the corresponding score thresholds obtained for each season could be used.

5.5 Evaluation of the final detection skill and recommendations

The detection skill of the final configuration of the MDI algorithm (see Table 4.5) is not as good as the detection skill of other methods. Additionally, it must be mentioned that this method, in its current state, can only be compared with existing work to a limited extent. First, up to now, the presented method has only dealt with the candidate search of TCs, the tracking of the corresponding storms, as it is done subsequently by most other methods, still needs to be developed. Second, each method is applied to different geographical regions. Some studies (e.g. Bengtsson et al. 1982; McDonald et al. 2005; Oouchi et al. 2006; Sugi et al. 2002) restrict the investigated region, others apply their detector to worldwide data. In this study, too, the investigated region and the examined period of time were limited, because otherwise the unsupervised algorithm acts too unspecific. It is therefore recommended not to understand the method as a classical detector for single TCs, but to indicate and measure local storm activity (see Section 4.5).
In order to better classify the detection skill of the MDI algorithm, the key figures of some other methods should be mentioned here: Tory et al. (2013a) state that they achieve a POD of 78 % with a false alarm rate of 25 %. Liu et al. (2016) state a POD of 98.9 % in detecting TCs in high-resolution models with a FAR of 0.03 %.
After closer visual inspection of the detections, the following observations can be made which shed more light on the detection skill presented in Table 4.5:

- As can be seen in Figure 4.13, the detections classified as false alarms are mostly at least tropical depressions or storms that are not labeled. Detections at the periphery of storms that do not cover the labeled storm center are also found (see Figure 4.13, panel "False alarms"). Sometimes the ground truth also passes through the detection box quickly, which means that the criterion that a detection box must be occupied by a ground truth at least during 25 % of its time steps is

not met. It is therefore rare that the MDI algorithm detects intervals that are not storm-like events.

- The misses, i.e. undetected ground truths, are often weak storms. All major hurricanes are detected.

- The high number of misses is partly also caused because storms are not detected at every time step, but mostly only once at the maximum of their activity in the examined area. This is visible, for instance, in Figure 4.12: First, the ground truths of Katrina are not detected, i.e. classified as miss. At the peak of Katrina's activity, the ground truths are finally covered by a detection box. Further north, the ground truths remain undetected again and are considered as misses. Effectively this storm was detected, but most of its track was outside the detection box, resulting in many misses.

- The BIAS is not meaningful in this application, because a hit is not equivalent to the detection of a storm, but to the coverage of a ground truth. In other words, in this study different things are compared with each other when calculating the BIAS (Equation 3.16): hits are the number of ground truths covered, the false alarms are the number of detection boxes that do not contain a ground truth. The misses are the number of uncovered ground truths. This means that the number of hits and misses is high, while the number of false alarms is low. A more meaningful measure would be a comparison of the total number of detection boxes (whether true or false detection) with the number of storms in this season. This, however, aims at the placement of one box per storm. Here, this measure would result in 1.25 (79/63), i.e. the detector reports slightly more storms than actually occur. It would make even more sense to place many detection boxes along the storm track, link them together and interpret them as one detection. For this the algorithm would have to be improved.
  Another problem is the fact that the detection boxes are not allowed to overlap each other. This can be seen for example in Figure 4.12. Around the maximum of Katrina, a detection box is placed, but further north no other box is placed. This is because the detection either got a low score and was discarded afterwards or because there is not enough space to place a detection box. In Section 4.2, a minimum size of 2 grid spacings was specified. If the boxes could overlap, in Katrina's example boxes would probably have been placed before and after the maximum, increasing the POD. However, if overlapping is allowed, this also complicates the evaluation: if two boxes overlap almost completely and cover a ground truth, it is questionable whether both boxes should be evaluated as correct or one of them as false alarm. This affects the measures of detection skill (see Section 3.4), as the number of false alarms will then vary greatly.

- The restriction that a ground truth must be contained for at least 25 % of the time steps of a detection could be discarded. This condition was originally introduced to allow tuning of the maximum duration of a detection box. Without this condition,

longer lasting boxes would have resulted in a better detection skill, as they cover a longer period in time and therefore contain more ground truths. However, since no such tuning was performed, but this length was fixed to 8 time steps, the condition is obsolete. In addition, the ground truths sometimes do not match exactly the storm center in the ERA-Interim data. As a result, it can happen that actually correct detections are classified as false alarms. Or the MDI algorithm detects peripheral areas of the storm, which is also acceptable. If the condition that 25 % of the time steps of a detection must contain a ground truth is removed, the POD increases to 0.474 and the PRC to 0.886.

5.5.1 Meaningful applications of the MDI algorithm

The utilization of the MDI algorithm as a precise detector for TCs, which could possibly be extended by a tracking algorithm in the future, is questionable in the light of the results. A possibility of how the algorithm can nevertheless be usefully applied to reanalysis and climate model data is described in Section 4.5. With little computational effort of a few seconds per hurricane season, the algorithm can estimate their (expected) strength using Equation 4.1. It is interesting to note that the MDI algorithm also returns detections when no TC activity is present. This is because even in calm seasons there are naturally anomalies present, even if these are not TCs.
In principle, it should be possible to apply the MDI algorithm to other regions as well, but detection skill would still have to be evaluated here. It should also be tested whether the detection skill is influenced by the model resolution. No high thresholds were used, but only soft ones, which have mainly the task to filter out detected calm intervals. Therefore, the dependence on the resolution should be small.
Furthermore, estimates of the development of future hurricane activity would be imaginable, for example, when applying the MDI algorithm to CMIP (Coupled Model Intercomparison Project) (Eyring et al. 2016) model data.

In addition, further interesting applications of MDI algorithm on climate data are imaginable. In general, this algorithm could be especially helpful for the investigation of anomalies, whose nature is less well known and defined than the one of TCs. The aim should not be the detection of exactly classified anomalies, but rather the first detection of anomalies that have not yet been explored. This is especially because the algorithm, as described by Barz et al. (2018), detects any kind of anomaly non-specifically: these can be pointwise or collective anomalies, small-scale or large-scale changes in amplitude or frequency of variables, contextual anomalies (such as low temperatures in summer) or change-point anomalies that divide a data set into several different domains, or a mixture of all of these. An application to long-term climate model data and the subsequent investigation of the high ranked anomalies would be a useful procedure. The MDI algorithm should therefore not be used to search for specific anomalies, whose general characteristics can be reasonably well described, but rather to shed light on unknown events in data sets that have an impact on humans and the environment.

6 | Conclusions and Outlook

In this study, the applicability of the MDI algorithm for the detection of TCs as well as its advantages and disadvantages compared to traditional detectors were investigated. For this purpose, the MDI algorithm was applied to different variable combinations of ERA-Interim reanalysis data of the hurricane seasons 2000-2010 in the Gulf of Mexico using different configuration settings. The IBTrACS data set, which contains observed historical TC locations and intensities, was used to evaluate the detections.

The following steps were carried out, which led to the conclusions mentioned as well:

- A case study was conducted to limit the choice of possible input variables, resulting in a selection of 7 variables at suitable pressure levels, namely relative vorticity, wind speed, geopotential height, absolute humidity, temperature and layer thickness.

- The optimal initial settings of the MDI algorithm were determined: The U-KL divergence proved to be the most suitable divergence method. Time-delay embedding did not turn out to be useful, so it was not used in the further study. In the case of spatial-neighbor embedding, a combination of $\kappa_{xy} = 2$ and $\tau_{xy} = 3$ performed best, followed closely by a completely switched-off embedding. With a readjustment by slightly shifting the detections one time step forward, which was referred to as time shift, the detection skill could be further improved, and the choice was finally made to not use any embedding.

- All possible variable combinations were examined in detail for their detection skill in all hurricane seasons considered. It turned out that the relative vorticity at 850 hPa is essential in the search for TCs. Each combination of variables performs better when vo850 is included and the mAP increases by a mAP of 0.277. The more variables are added to vo850, the worse the detection skill. Additionally, TCs cause detectable anomalies in the absolute humidity at 700 hPa and the relative vorticity at 250 hPa. The application to variables associated with the warm core, specifically the temperature at 400 hPa and the layer thickness between 300 and 500 hPa reduces the detection skill.

- The detection skill could be further improved by post-processing: by filtering out false alarms using soft thresholds, the mAP was increased by about 0.02. A threshold for wind speed of $8\,\mathrm{ms}^{-1}$ and relative vorticity of $1 \cdot 10^{-5}\,\mathrm{s}^{-1}$ was chosen.

With these low thresholds it was ensured that only calm intervals were discarded. Thus, the MDI algorithm could also be used at other resolutions where TCs. By means of a cutoff threshold, which excludes all detections with a score below 1150, a precision of 0.838 and a false alarm rate of 0.162 was achieved. With a POD of 0.455 almost half of all ground truths are detected. This mediocre detection skill is put into perspective by considering that the MDI algorithm currently does not track TCs, but usually only detects them at their maximum. This means that part of the track respectively part of the ground truths is not detected.

- The MDI algorithm can be used to assess the strength of a hurricane season in a specific area. A statistically highly significant linear correlation ($r = 0.9$) between the sum of the detection scores and the ACE was found.

In summary, the MDI algorithm is currently suitable for the detection of TCs at their maximum. However, there is still development potential to further optimize the MDI algorithm as TC detector:

- With the current configuration, tracking of TCs is not yet possible. This would require further research and the MDI algorithm could possibly be extended by a tracking algorithm.
- It should be investigated how the MDI algorithm behaves as a TC detector when applied to other regions and other temporal and spatial resolutions.
- The MDI algorithm should be applied to larger data sets and longer time series in order to be able to compare its detection skills with other TC detectors.
- An attempt should be made to apply the MDI algorithm not only to individual pressure levels, but spatially in three dimensions.

In the final assessment of the MDI algorithm as a TC detector, the following should be noted: The MDI algorithm acts more unspecific in the search for TC candidates than traditional methods, which is why mainly the storm maximum is detected and TCs in their initial phase or weak TCs are not detected. The false alarm rate is low compared to e.g. Tory et al. (2013a). The advantages over conventional methods is the absence of hard thresholds of variables, which makes it better suited for application to other resolutions. However, the latter thresholds are replaced by the score threshold, which is more difficult to understand and to adjust for users who are not familiar with the MDI algorithm.
Regarding the current state of research, it was shown how weather extremes can be detected with a novel approach without resorting to the traditional threshold methods. This shows that unsupervised machine learning algorithms can work as detectors for extreme weather events without the assistance of other algorithms or interaction with the user. Application of MDI algorithm is more helpful in detecting unknown or barely defined anomalies. In addition, MDI algorithm scores can be useful in assessing the extent to which certain events deviate from climatological conditions. In this way, individual events could be better assessed in the context of climate change.

Bibliography

Adler, R. (2005). "Estimating the benefit of TRMM tropical cyclone data in saving lives. American meteorological society". In: *15th Conference on Applied Climatology, Savannah*.

Barz, B., Y. G. Garcia, E. Rodner, and J. Denzler (2017). "Maximally divergent intervals for extreme weather event detection". In: *OCEANS 2017, Aberdeen*. IEEE, pp. 1–9.

Barz, B., E. Rodner, Y. G. Garcia, and J. Denzler (2018). "Detecting regions of maximal divergence for spatio-temporal anomaly detection". In: *IEEE transactions on pattern analysis and machine intelligence*.

Bell, G. D., M. S. Halpert, R. C. Schnell, R. W. Higgins, J. Lawrimore, V. E. Kousky, R. Tinker, W. Thiaw, M. Chelliah, and A. Artusa (2000). "Climate assessment for 1999". In: *Bulletin of the American Meteorological Society* 81.6, pp. 1–50.

Bengtsson, L., H. Böttger, and M. Kanamitsu (1982). "Simulation of hurricane-type vortices in a general circulation model". In: *Tellus* 34.5, pp. 440–457.

Bengtsson, L., M. Botzet, and M. Esch (1995). "Hurricane-type vortices in a general circulation model". In: *Tellus A* 47.2, pp. 175–196.

– (1996). "Will greenhouse gas-induced warming over the next 50 years lead to higher frequency and greater intensity of hurricanes?" In: *Tellus A* 48.1, pp. 57–73.

Bengtsson, L., K. Hodges, and M. Esch (2007). "Hurricane type vortices in a high-resolution global model: Comparison with observations and reanalyses". In: *Tellus*.

Bister, M. and K. A. Emanuel (1997). "The genesis of Hurricane Guillermo: TEXMEX analyses and a modeling study". In: *Monthly Weather Review* 125.10, pp. 2662–2682.

Briegel, L. M. and W. M. Frank (1997). "Large-scale influences on tropical cyclogenesis in the western North Pacific". In: *Monthly Weather Review* 125.7, pp. 1397–1413.

Broccoli, A. and S. Manabe (1990). "Can existing climate models be used to study anthropogenic changes in tropical cyclone climate?" In: *Geophysical Research Letters* 17.11, pp. 1917–1920.

Camargo, S. J. and S. E. Zebiak (2002). "Improving the detection and tracking of tropical cyclones in atmospheric general circulation models". In: *Weather and forecasting* 17.6, pp. 1152–1162.

Caron, L.-P., C. G. Jones, P. A. Vaillancourt, and K. Winger (2013). "On the relationship between cloud–radiation interaction, atmospheric stability and Atlantic tropical cyclones in a variable-resolution climate model". In: *Climate Dynamics* 40.5-6, pp. 1257–1269.

Caron, L.-P., C. G. Jones, and K. Winger (2011). "Impact of resolution and downscaling technique in simulating recent Atlantic tropical cyclone activity". In: *Climate dynamics* 37.5-6, pp. 869–892.

Chan, J. C. (2010). "Movement of tropical cyclones". In: *Global perspectives on Tropical cyclones: From science to mitigation*. World Scientific, pp. 133–148.

Chandola, V., A. Banerjee, and V. Kumar (2009). "Anomaly detection: A survey". In: *ACM computing surveys (CSUR)* 41.3, p. 15.

Chawla, S. and P. Sun (2006). "SLOM: a new measure for local spatial outliers". In: *Knowledge and Information Systems* 9.4, pp. 412–429.

Cheng, T. and Z. Li (2006). "A multiscale approach for spatio-temporal outlier detection". In: *Transactions in GIS* 10.2, pp. 253–263.

Davis, C. A. and L. F. Bosart (2003). "Baroclinically induced tropical cyclogenesis". In: *Monthly Weather Review* 131.11, pp. 2730–2747.

Dee, D. P., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, et al. (2011). "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system". In: *Quarterly Journal of the royal meteorological society* 137.656, pp. 553–597.

DeMaria, M. and J. Kaplan (1994). "A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin". In: *Weather and Forecasting* 9.2, pp. 209–220.

– (1999). "An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins". In: *Weather and Forecasting* 14.3, pp. 326–337.

DeMaria, M., M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan (2005). "Further improvements to the statistical hurricane intensity prediction scheme (SHIPS)". In: *Weather and Forecasting* 20.4, pp. 531–543.

Demšar, J. (2006). "Statistical comparisons of classifiers over multiple data sets". In: *Journal of Machine learning research*, pp. 1–30.

Dunkerton, T. J., M. T. Montgomery, and Z. Wang (2008). "Tropical cyclogenesis in a tropical wave critical layer: Easterly waves". In: *Atmospheric Chemistry and Physics Discussions* 8.3, pp. 11149–11292.

Eliassen, A. (1951). "Slow thermally or frictionally controlled meridional circulation in a circular vortex". In: *Astrophysica Norvegica* 5, p. 19.

Evans, C. and R. E. Hart (2008). "Analysis of the wind field evolution associated with the extratropical transition of Bonnie (1998)". In: *Monthly Weather Review* 136.6, pp. 2047–2065.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor (2016). "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization". In: *Geoscientific Model Development* 9.

Fiorino, M. and R. L. Elsberry (1989). "Some aspects of vortex structure related to tropical cyclone motion". In: *Journal of the atmospheric sciences* 46.7, pp. 975–990.

Frank, N. L. and G. Clark (1980). "Atlantic tropical systems of 1979". In: *Monthly Weather Review* 108.7, pp. 966–972.

Frank, W. M. and E. A. Ritchie (1999). "Effects of environmental flow upon tropical cyclone structure". In: *Monthly Weather Review* 127.9, pp. 2044–2061.

– (2001). "Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes". In: *Monthly Weather Review* 129.9, pp. 2249–2269.

Frank, W. M. and G. S. Young (2007). "The interannual variability of tropical cyclones". In: *Monthly Weather Review* 135.10, pp. 3587–3598.

Franklin, J. L., M. L. Black, and K. Valde (2003). "GPS dropwindsonde wind profiles in hurricanes and their operational implications". In: *Weather and Forecasting* 18.1, pp. 32–44.

Friedman, M. (1937). "The use of ranks to avoid the assumption of normality implicit in the analysis of variance". In: *Journal of the American Statistical Association* 32.200, pp. 675–701.

– (1940). "A comparison of alternative tests of significance for the problem of m rankings". In: *The Annals of Mathematical Statistics* 11.1, pp. 86–92.

Gallina, G. M. (2002). *Environmental vertical wind shear and tropical cyclone intensity change utilizing enhanced satellite derived wind information*. University of Wisconsin–Madison.

Gray, W. M. (1967). "Global view of the origin of tropical disturbances and storms". In: *Atmospheric science paper* 114.

– (1979). "Hurricanes: their formation, structure and likely role in the tropical circulation. Meteorology over the tropical oceans". In: *Quarterly Journal of the Royal Meteorological Society*, pp. 155–218.

– (1985). "Tropical cyclone global climatology". In: *WMO Technical Document WMO/TD* 72.1, pp. 3–19.

Haarsma, R. J., J. F. Mitchell, and C. Senior (1993). "Tropical disturbances in a GCM". In: *Climate Dynamics* 8.5, pp. 247–257.

Harr, P. A. (2010). "The extratropical transition of tropical cyclones: Structural characteristics, downstream impacts, and forecast challenges". In: *Global Perspectives on Tropical Cyclones: From Science to Mitigation*. World Scientific, pp. 149–174.

Haurwitz, B. (1935). *The height of tropical cyclones and the "eye" of the storm*. US Government Printing Office.

Hawkins, D. M. (1980). *Identification of outliers*. Vol. 11. Springer.

Hendricks, E., M. T. Montgomery, and C. Davis (2004). "On the role of "vortical" hot towers in hurricane formation". In: *Journal of the Atmospheric Sciences* 61, pp. 1209–1232.

Hodges, K., A. Cobb, and P. L. Vidale (2017). "How well are tropical cyclones represented in reanalysis datasets?" In: *Journal of Climate* 30.14, pp. 5243–5264.

Horn, M., K. Walsh, M. Zhao, S. J. Camargo, E. Scoccimarro, H. Murakami, H. Wang, A. Ballinger, A. Kumar, D. A. Shaevitz, et al. (2014). "Tracking scheme dependence

of simulated tropical cyclone response to idealized climate simulations". In: *Journal of climate* 27.24, pp. 9197–9213.

Hotelling, H. (1947). "Multivariate quality control". In: *Techniques of statistical analysis.*

Houze Jr., R. A. (2004). "Mesoscale convective systems". In: *Reviews of Geophysics* 42.4.

Intergovernmental Panel on Climate Change (IPCC), T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley, et al. (2013). *Climate change 2013: The physical science basis.*

Keogh, E., J. Lin, and A. Fu (2005). "Hot sax: Efficiently finding the most unusual time series subsequence". In: *Fifth IEEE International Conference on Data Mining (ICDM'05)*. IEEE, p. 8.

Kepert, J. D. (2006a). "Observed boundary layer wind structure and balance in the hurricane core. Part I: Hurricane Georges". In: *Journal of the atmospheric sciences* 63.9, pp. 2169–2193.

– (2006b). "Observed boundary layer wind structure and balance in the hurricane core. Part II: Hurricane Mitch". In: *Journal of the atmospheric sciences* 63.9, pp. 2194–2211.

– (2010). "Tropical cyclone structure and dynamics". In: *Global perspectives on Tropical cyclones: from science to mitigation*. World Scientific, pp. 3–53.

King, D., J. Davidson, and L. Anderson-Berry (2010). "Disaster mitigation and societal impacts". In: *Global Perspectives on Tropical Cyclones: From Science to Mitigation*. World Scientific, pp. 409–436.

Klein, P. M., P. A. Harr, and R. L. Elsberry (2000). "Extratropical transition of western North Pacific tropical cyclones: An overview and conceptual model of the transformation stage". In: *Weather and Forecasting* 15.4, pp. 373–395.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann (2010). "The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data". In: *Bulletin of the American Meteorological Society* 91.3, pp. 363–376.

Knutson, T. R., J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. Srivastava, and M. Sugi (2010). "Tropical cyclones and climate change". In: *Nature geoscience* 3.3, p. 157.

Knutson, T. R., J. J. Sirutis, S. T. Garner, I. M. Held, and R. E. Tuleya (2007). "Simulation of the recent multidecadal increase of Atlantic hurricane activity using an 18-km-grid regional model". In: *Bulletin of the American Meteorological Society* 88.10, pp. 1549–1565.

Kotsiantis, S. B., I. Zaharakis, and P. Pintelas (2007). "Supervised machine learning: A review of classification techniques". In: *Emerging artificial intelligence applications in computer engineering* 160, pp. 3–24.

Kurihara, Y. and R. E. Tuleya (1981). "A numerical simulation study on the genesis of a tropical storm". In: *Monthly Weather Review* 109.8, pp. 1629–1653.

Laing, A. and J. Evans (2011). *An Introduction to Tropical Meteorology, The COMET Program.*

Liu, S., M. Yamada, N. Collier, and M. Sugiyama (2013). "Change-point detection in time-series data by relative density-ratio estimation". In: *Neural Networks* 43, pp. 72–83.

Liu, Y., E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, et al. (2016). "Application of deep convolutional neural networks for detecting extreme weather in climate datasets". In: *ACM SIGKDD 2016: Conference on Knowledge Discovery and Data Mining*.

Maynard, R. H. (1945). "Radar and weather". In: *Journal of Meteorology* 2.4, pp. 214–226.

McBride, J. L. and R. Zehr (1981). "Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems". In: *Journal of the Atmospheric Sciences* 38.6, pp. 1132–1151.

McDonald, R. E., D. G. Bleaken, D. R. Cresswell, V. D. Pope, and C. A. Senior (2005). "Tropical storms: representation and diagnosis in climate models and the impacts of climate change". In: *Climate Dynamics* 25.1, pp. 19–36.

McTaggart-Cowan, R., G. D. Deane, L. F. Bosart, C. A. Davis, and T. J. Galarneau Jr (2008). "Climatology of tropical cyclogenesis in the North Atlantic (1948–2004)". In: *Monthly Weather Review* 136.4, pp. 1284–1304.

Merrill, R. T. (1984). "A comparison of large and small tropical cyclones". In: *Monthly Weather Review* 112.7, pp. 1408–1418.

Montgomery, M. T., M. Nicholls, T. Cram, and A. Saunders (2006). "A vortical hot tower route to tropical cyclogenesis". In: *Journal of the atmospheric sciences* 63.1, pp. 355–386.

Munich RE (2019). *NatCatSERVICE: natural catastrophe know-how for risk management and research*. Tech. rep. Munich RE.

Murakami, H. (2014). "Tropical cyclones in reanalysis data sets". In: *Geophysical Research Letters* 41.6, pp. 2133–2141.

Murakami, H. and M. Sugi (2010). "Effect of model resolution on tropical cyclone climate projections". In: *Sola* 6, pp. 73–76.

Murakami, H., Y. Wang, H. Yoshimura, R. Mizuta, M. Sugi, E. Shindo, Y. Adachi, S. Yukimoto, M. Hosaka, S. Kusunoki, et al. (2012). "Future changes in tropical cyclone activity projected by the new high-resolution MRI-AGCM". In: *Journal of Climate* 25.9, pp. 3237–3260.

National Centers for Environmental Information (NCEI) (2017). *An Inventory of Tropical Cyclone Tracks*. URL: `https://www.ncei.noaa.gov/news/inventory-tropical-cyclone-tracks` (visited on 10/16/2019).

Nemenyi, P. (1963). "Distribution-free multiple comparisons". dissertation. Princeton University.

National Hurricane Center (NHC) (2012). *2012 Revision to the Saffir-Simpson Hurricane Wind Scale*. URL: `https://www.nhc.noaa.gov/pdf/sshws_2012rev.pdf` (visited on 07/29/2019).

Oouchi, K., J. Yoshimura, H. Yoshimura, R. Mizuta, S. Kusunoki, and A. Noda (2006). "Tropical cyclone climatology in a global-warming climate as simulated in a 20 km-mesh global atmospheric model: Frequency and wind intensity analyses". In: *Journal of the Meteorological Society of Japan. Ser. II* 84.2, pp. 259–276.

Ooyama, K. (1969). "Numerical simulation of the life cycle of tropical cyclones". In: *Journal of the Atmospheric Sciences* 26.1, pp. 3–40.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw (1980). "Geometry from a time series". In: *Physical review letters* 45.9, p. 712.

Pielke Jr., R. A., J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin (2008). "Normalized hurricane damage in the United States: 1900–2005". In: *Natural Hazards Review* 9.1, pp. 29–42.

Powell, M. D., P. J. Vickery, and T. A. Reinhold (2003). "Reduced drag coefficient for high wind speeds in tropical cyclones". In: *Nature* 422.6929, p. 279.

Ren, H., M. Liu, X. Liao, L. Liang, Z. Ye, and Z. Li (2018). "Anomaly detection in time series based on interval sets". In: *IEEJ Transactions on Electrical and Electronic Engineering* 13.5, pp. 757–762.

Rotunno, R. and K. A. Emanuel (1987). "An air–sea interaction theory for tropical cyclones. Part II: Evolutionary study using a nonhydrostatic axisymmetric numerical model". In: *Journal of the Atmospheric Sciences* 44.3, pp. 542–561.

Roundy, P. E. and W. M. Frank (2004). "A climatology of waves in the equatorial region". In: *Journal of the atmospheric sciences* 61.17, pp. 2105–2132.

Saito, T. and M. Rehmsmeier (2015). "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets". In: *PloS one* 10.3.

Schenkel, B. A. and R. E. Hart (2012). "An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets". In: *Journal of Climate* 25.10, pp. 3453–3475.

Shannon, C. E. (1948). "A mathematical theory of communication". In: *Bell system technical journal* 27.3, pp. 379–423.

Simpson, J., E. Ritchie, G. Holland, J. Halverson, and S. Stewart (1997). "Mesoscale interactions in tropical cyclone genesis". In: *Monthly Weather Review* 125.10, pp. 2643–2661.

Simpson, R. H. and H. Saffir (1974). "The hurricane disaster potential scale". In: *Weatherwise* 27.8, p. 169.

Smith, R. K. and M. T. Montgomery (2016). "Understanding hurricanes". In: *Weather* 71.9, pp. 219–223.

Smith, R. K. and S. Vogl (2008). "A simple model of the hurricane boundary layer revisited". In: *Quarterly Journal of the Royal Meteorological Society* 134.631, pp. 337–351.

Strachan, J., P. L. Vidale, K. Hodges, M. Roberts, and M.-E. Demory (2013). "Investigating global tropical cyclone activity with a hierarchy of AGCMs: The role of model resolution". In: *Journal of Climate* 26.1, pp. 133–152.

Sugi, M., A. Noda, and N. Sato (2002). "Influence of the global warming on tropical cyclone climatology: An experiment with the JMA global model". In: *Journal of the Meteorological Society of Japan. Ser. II* 80.2, pp. 249–272.

Takens, F. (1981). "Detecting strange attractors in turbulence". In: *Dynamical systems and turbulence, Warwick 1980*. Springer, pp. 366–381.

Tang, C. and C. Monteleoni (2015). "Can topic modeling shed light on climate extremes?" In: *Computing in Science & Engineering*.

Thorncroft, C. and S. C. Jones (2000). "The extratropical transitions of Hurricanes Felix and Iris in 1995". In: *Monthly Weather Review* 128.4, pp. 947–972.

Tory, K. J., S. Chand, R. Dare, and J. McBride (2013a). "The development and assessment of a model-, grid-, and basin-independent tropical cyclone detection scheme". In: *Journal of Climate* 26.15, pp. 5493–5507.

Tory, K. J. and W. M. Frank (2010). "Tropical cyclone formation". In: *Global perspectives on tropical cyclones: From science to mitigation*. World Scientific, pp. 55–91.

Tory, K., R. Dare, N. Davidson, J. McBride, and S. Chand (2013b). "The importance of low-deformation vorticity in tropical cyclone formation". In: *Atmospheric Chemistry and Physics* 13.4, pp. 2115–2132.

Tsutsui, J.-i. and A. Kasahara (1996). "Simulated tropical cyclones using the National Center for Atmospheric Research community climate model". In: *Journal of Geophysical Research: Atmospheres* 101.D10, pp. 15013–15032.

Ullrich, P. A. and C. M. Zarzycki (2016). "TempestExtremes v1. 0: A framework for scale-insensitive pointwise feature tracking on unstructured grids". In: *Geoscientific Model Development*.

Vitart, F., J. Anderson, J. Sirutis, and R. Tuleya (2001). "Sensitivity of tropical storms simulated by a general circulation model to changes in cumulus parametrization". In: *Quarterly Journal of the Royal Meteorological Society* 127.571, pp. 25–51.

Vitart, F., J. Anderson, and W. Stern (1997). "Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations". In: *Journal of Climate* 10.4, pp. 745–760.

– (1999). "Impact of large-scale circulation on tropical storm frequency, intensity, and location, simulated by an ensemble of GCM integrations". In: *Journal of climate* 12.11, pp. 3237–3254.

Vitart, F., D. Anderson, and T. Stockdale (2003). "Seasonal forecasting of tropical cyclone landfall over Mozambique". In: *Journal of Climate* 16.23, pp. 3932–3945.

Walsh, K. J. (1997). "Objective detection of tropical cyclones in high-resolution analyses". In: *Monthly Weather Review* 125.8, pp. 1767–1779.

Walsh, K. J., M. Fiorino, C. Landsea, and K. McInnes (2007). "Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses". In: *Journal of climate* 20.10, pp. 2307–2314.

Walsh, K. J. and J. J. Katzfey (2000). "The impact of climate change on the poleward movement of tropical cyclone–like vortices in a regional climate model". In: *Journal of Climate* 13.6, pp. 1116–1132.

Walsh, K. J. and I. G. Watterson (1997). "Tropical cyclone-like vortices in a limited area model: comparison with observed climatology". In: *Journal of climate* 10.9, pp. 2240–2259.

Walsh, K. J., J. L. McBride, P. J. Klotzbach, S. Balachandran, S. J. Camargo, G. Holland, T. R. Knutson, J. P. Kossin, T.-c. Lee, A. Sobel, et al. (2016). "Tropical cyclones and climate change". In: *Wiley Interdisciplinary Reviews: Climate Change* 7.1, pp. 65–89.

Wang, Y. and G. J. Holland (1996). "Tropical cyclone motion and evolution in vertical shear". In: *Journal of the atmospheric sciences* 53.22, pp. 3313–3332.

Wexler, H. (1947). "Structure of hurricanes as determined by radar". In: *Annals of the New York Academy of Sciences* 48.8, pp. 821–845.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Vol. 100. Academic press.

Willoughby, H., J. Clos, and M. Shoreibah (1982). "Concentric eye walls, secondary wind maxima, and the evolution of the hurricane vortex". In: *Journal of the Atmospheric Sciences* 39.2, pp. 395–411.

Wong, M. L. and J. C. Chan (2004). "Tropical cyclone intensity in vertical wind shear". In: *Journal of the atmospheric sciences* 61.15, pp. 1859–1876.

Wu, E., W. Liu, and S. Chawla (2008). "Spatio-temporal outlier detection in precipitation data". In: *International Workshop on Knowledge Discovery from Sensor Data*. Springer, pp. 115–133.

Wu, G. and N.-C. Lau (1992). "A GCM simulation of the relationship between tropical-storm formation and ENSO". In: *Monthly Weather Review* 120.6, pp. 958–977.

Zarzycki, C. M. and P. A. Ullrich (2017). "Assessing sensitivities in algorithmic detection of tropical cyclones in climate data". In: *Geophysical Research Letters* 44.2, pp. 1141–1149.

Zehr, R. M. (1992). "Tropical cyclogenesis in the western North Pacific". In:

Zhao, M., I. M. Held, S.-J. Lin, and G. A. Vecchi (2009). "Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM". In: *Journal of Climate* 22.24, pp. 6653–6678.

Front page figure: Hurricane Katrina on 28 August 2005 at 18:00 UTC as it appears in ERA-interim data. Clockwise starting at the top left panel, the variables relative vorticity at 850 hPa, geopotential height at 1000 hPa, temperature at 400 hPa and absolute humidity at 700 hPa are shown. Bluish colors indicate low values, greenish medium and reddish high values relative to the whole summer season.

Acknowledgments

First, I would like to thank Prof. Dr. habil. Veronika Eyring for the opportunity to write this thesis in the department "Earth System Model Evaluation and Analysis" of the DLR in Oberpfaffenhofen. Furthermore, I would like to thank her for the interesting topic proposal, excellent scientific supervision and advice regarding my future career. I would also like to thank Prof. Dr. Markus Rapp for his sympathetic ear and for his friendly and helpful cooperation. I thank Björn Barz for his broad assistance in handling the algorithm he developed. I thank Prof. Dr. Joachim Denzler for the interesting suggestions from the perspective of the Computer Vision Group of the University of Jena. This also applies to all the other members of this research group whom I had the pleasure to meet in Jena. I would like to thank Axel Lauer for his extensive and detailed comments on my work. I am grateful to Björn Brötz for the introduction to the world of the shell with all its little tricks and for musical inspiration. Furthermore, I would like to express my thanks to my office colleagues Kévin, Gunnar and Manuel, who not only gave excellent remarks on the subject, but were also good hiking companions and helped me to improve my foreign language skills. I would like to thank Dr. Birgit Hassler for a wonderful first day at DLR and for many more delightful conversations about Arctic regions. As excellent companions in the escape room, I would like to thank Dr. Lisa Bock and Michaela Langer, as well as the other members of our working group and the Institute of Atmospheric Physics, who contributed to a great time at work. Personally, I would like to thank my wonderful parents, who always support me in everything I strive for and who are a constant help in the challenging times of life. I also want to thank my sisters and my entire family for their unconditional love and incomparable solidarity. Finally, I would like to thank Luisa, with whom I have enjoyed and mastered beautiful and stressful phases of the recent year and who inspires me like no one else does.

A | Appendix

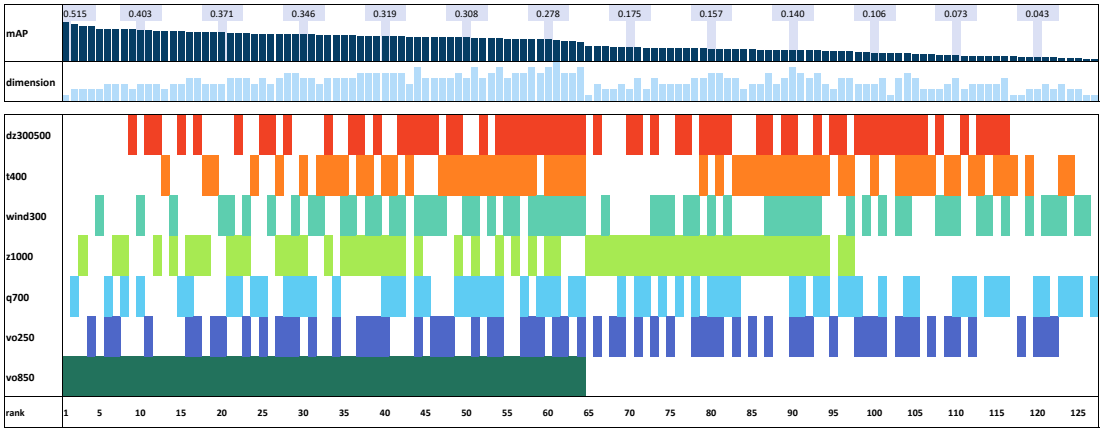


FIGURE A.1: All 127 variable combinations ranked according to their mAP, which is obtained by the detection of TCs. Only spatial-neighbor embedding was applied, so the embedding parameters (1,1,2,3) were used and a time shift of $\Delta t = -1$ was chosen. On the left one finds the best performing combination, indicated by rank 1, on the right the worst performing combination. The color code above the ranks indicates which variables have been selected for this run. In addition, the dimension is indicated by bars, i.e. whether the MDI algorithm was applied univariate (dimension=1) or multivariate to up to seven variables (dimension=7). Above this, the mAP of the combination, i.e. the mean value of the APs from all hurricane seasons, is additionally indicated by bars. Every 10^{th} rank the mAP is mentioned by its value for orientation.

TABLE A.1: Complete tables of the Nemenyi tests. The upper table shows the p-values when comparing different time-delay embedding parameters in pairs, the lower left table between the combinations of spatial-neighbor and time-delay embedding parameters and the lower right table between different time shifts. The p-values colored turquoise are significant, green colored ones are highly significant.

Time-delay parameter

κτ	11	21	31	41	51	61	22	32	42	52	62	23	33	43	53	63	24	34	44	54	64	25	35	45	55	65	26	36	46	56	66
11		0,900	0,900	0,680	0,001	0,001	0,900	0,122	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,897	0,001	0,001	0,001	0,001	0,024	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
21	0,900		0,900	0,815	0,003	0,001	0,900	0,208	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,049	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
31	0,900	0,900		0,900	0,186	0,025	0,900	0,900	0,003	0,001	0,001	0,900	0,036	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,659	0,001	0,001	0,001	0,001	0,004	0,001	0,001	0,001	0,001
41	0,680	0,815	0,900		0,900	0,633	0,900	0,900	0,269	0,001	0,001	0,900	0,701	0,001	0,001	0,001	0,900	0,004	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,310	0,001	0,001	0,001	0,001
51	0,001	0,003	0,186	0,900		0,900	0,035	0,900	0,900	0,003	0,001	0,027	0,900	0,001	0,001	0,001	0,841	0,833	0,001	0,001	0,001	0,900	0,659	0,001	0,001	0,001	0,900	0,054	0,001	0,001	0,001
61	0,001	0,001	0,025	0,633	0,900		0,003	0,900	0,900	0,038	0,001	0,002	0,900	0,001	0,001	0,001	0,406	0,900	0,001	0,001	0,001	0,900	0,900	0,001	0,001	0,001	0,900	0,312	0,001	0,001	0,001
22	0,900	0,900	0,900	0,900	0,035	0,003		0,622	0,001	0,001	0,001	0,900	0,004	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,274	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
32	0,122	0,208	0,900	0,900	0,900	0,900	0,622		0,886	0,001	0,001	0,574	0,900	0,001	0,001	0,001	0,900	0,098	0,001	0,001	0,001	0,900	0,042	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001
42	0,001	0,001	0,003	0,269	0,900	0,900	0,001	0,886		0,184	0,001	0,001	0,900	0,001	0,001	0,001	0,116	0,900	0,001	0,001	0,001	0,900	0,900	0,001	0,001	0,001	0,900	0,677	0,001	0,001	0,001
52	0,001	0,001	0,001	0,001	0,003	0,038	0,001	0,001	0,184		0,001	0,001	0,027	0,900	0,276	0,015	0,001	0,900	0,516	0,172	0,123	0,001	0,900	0,900	0,685	0,276	0,156	0,900	0,900	0,100	0,346
62	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001		0,001	0,001	0,071	0,900	0,900	0,001	0,001	0,799	0,900	0,900	0,001	0,001	0,247	0,630	0,900	0,001	0,001	0,085	0,900	0,900
23	0,900	0,900	0,900	0,900	0,027	0,002	0,900	0,574	0,001	0,001	0,001		0,003	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,232	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
33	0,001	0,001	0,036	0,701	0,900	0,900	0,004	0,900	0,900	0,027	0,001	0,003		0,001	0,001	0,001	0,484	0,900	0,001	0,001	0,001	0,900	0,900	0,001	0,001	0,001	0,900	0,247	0,001	0,001	0,001
43	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,900	0,071	0,001	0,001		0,900	0,778	0,001	0,061	0,900	0,900	0,900	0,001	0,134	0,900	0,900	0,900	0,001	0,860	0,900	0,900	0,900
53	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,276	0,900	0,001	0,001	0,900		0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,032	0,900	0,900	0,900
63	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,015	0,900	0,001	0,001	0,778	0,900		0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,817	0,900	0,900
24	0,897	0,900	0,900	0,900	0,841	0,406	0,900	0,900	0,116	0,001	0,001	0,900	0,484	0,001	0,001	0,001		0,001	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,001	0,137	0,001	0,001	0,001	0,001
34	0,001	0,001	0,001	0,004	0,833	0,900	0,001	0,098	0,900	0,900	0,001	0,001	0,900	0,061	0,001	0,001	0,001		0,001	0,001	0,001	0,343	0,900	0,012	0,001	0,001	0,900	0,900	0,050	0,001	0,001
44	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,516	0,799	0,001	0,001	0,900	0,900	0,900	0,001	0,001		0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,093	0,900	0,900	0,900
54	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,172	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900		0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,015	0,900	0,900	0,900
64	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,123	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900		0,001	0,001	0,900	0,900	0,900	0,001	0,009	0,900	0,900	0,900
25	0,024	0,049	0,659	0,900	0,900	0,900	0,274	0,900	0,900	0,001	0,001	0,232	0,900	0,001	0,001	0,001	0,900	0,343	0,001	0,001	0,001		0,186	0,001	0,001	0,001	0,900	0,004	0,001	0,001	0,001
35	0,001	0,001	0,001	0,001	0,659	0,900	0,001	0,042	0,900	0,900	0,001	0,001	0,900	0,134	0,001	0,001	0,001	0,900	0,001	0,001	0,001	0,186		0,032	0,004	0,001	0,900	0,900	0,115	0,001	0,001
45	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,900	0,247	0,001	0,001	0,900	0,900	0,900	0,001	0,012	0,900	0,900	0,900	0,001	0,032		0,900	0,900	0,001	0,559	0,900	0,900	0,900
55	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,685	0,630	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,004	0,900		0,900	0,001	0,190	0,900	0,900	0,900
65	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,276	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900		0,001	0,032	0,900	0,900	0,900
26	0,001	0,001	0,004	0,310	0,900	0,900	0,001	0,900	0,900	0,156	0,001	0,001	0,900	0,001	0,001	0,001	0,137	0,900	0,001	0,001	0,001	0,900	0,900	0,001	0,001	0,001		0,635	0,001	0,001	0,001
36	0,001	0,001	0,001	0,001	0,054	0,312	0,001	0,001	0,677	0,900	0,001	0,001	0,247	0,860	0,032	0,001	0,001	0,900	0,093	0,015	0,009	0,004	0,900	0,559	0,190	0,032	0,635		0,820	0,007	0,045
46	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,900	0,085	0,001	0,001	0,900	0,900	0,817	0,001	0,050	0,900	0,900	0,900	0,001	0,115	0,900	0,900	0,900	0,001	0,820		0,900	0,900
56	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,100	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,007	0,900		0,900
66	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,346	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,001	0,900	0,900	0,900	0,001	0,045	0,900	0,900	

Time-delay and spatial-neighbor parameter

κ τ κ_{xy} τ_{xy}	1111	1121	1122	1123	1125	2111	2121	2123	2125	2311	2321	2323	2325	2511	2521	2523	2525
1111		0,9	0,001	0,001	0,9	0,9	0,9	0,266	0,156	0,652	0,9	0,858	0,707	0,001	0,001	0,199	0,001
1121	0,9		0,001	0,001	0,9	0,52	0,9	0,762	0,015	0,179	0,9	0,9	0,22	0,001	0,001	0,021	0,001
1122	0,001	0,001		0,9	0,001	0,001	0,001	0,123	0,001	0,001	0,001	0,006	0,001	0,001	0,001	0,001	0,001
1123	0,001	0,001	0,9		0,001	0,001	0,001	0,162	0,001	0,001	0,001	0,009	0,001	0,001	0,001	0,001	0,001
1125	0,9	0,9	0,001	0,001		0,516	0,9	0,766	0,015	0,176	0,9	0,9	0,216	0,001	0,001	0,021	0,001
2111	0,9	0,52	0,001	0,001	0,516		0,803	0,001	0,9	0,9	0,9	0,023	0,9	0,001	0,017	0,9	0,001
2121	0,9	0,9	0,001	0,001	0,9	0,803		0,478	0,065	0,449	0,9	0,9	0,506	0,001	0,001	0,087	0,001
2123	0,266	0,762	0,123	0,162	0,766	0,001	0,478		0,001	0,001	0,3	0,9	0,001	0,001	0,001	0,001	0,001
2125	0,156	0,015	0,001	0,001	0,015	0,9	0,065	0,001		0,9	0,134	0,001	0,9	0,048	0,543	0,9	0,001
2311	0,652	0,179	0,001	0,001	0,176	0,9	0,449	0,001	0,9		0,616	0,003	0,9	0,002	0,099	0,9	0,001
2321	0,9	0,9	0,001	0,001	0,9	0,9	0,9	0,3	0,134	0,616		0,894	0,671	0,001	0,001	0,174	0,001
2323	0,858	0,9	0,006	0,009	0,9	0,023	0,9	0,9	0,001	0,003	0,894		0,004	0,001	0,001	0,001	0,001
2325	0,707	0,22	0,001	0,001	0,216	0,9	0,506	0,001	0,9	0,9	0,671	0,004		0,002	0,078	0,9	0,001
2511	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,048	0,002	0,001	0,001	0,002		0,9	0,035	0,9
2521	0,001	0,001	0,001	0,001	0,001	0,017	0,001	0,001	0,543	0,099	0,001	0,001	0,078	0,9		0,478	0,830
2523	0,199	0,021	0,001	0,001	0,021	0,9	0,087	0,001	0,9	0,9	0,174	0,001	0,9	0,035	0,478		0,001
2525	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,9	0,83	0,001	

Time shift

κ τ κ_{xy} τ_{xy} Δt	1111 0	1111 -1	1111 -2	1111 -3	1111 -4	1123 0	1123 -1	1123 -2	1123 -3	1123 -4
1111 0		0,900	0,900	0,024	0,001	0,001	0,001	0,081	0,900	0,001
1111 -1	0,900		0,900	0,001	0,001	0,001	0,001	0,544	0,900	0,001
1111 -2	0,900	0,900		0,103	0,001	0,001	0,001	0,017	0,900	0,001
1111 -3	0,024	0,001	0,103		0,008	0,001	0,001	0,001	0,030	0,900
1111 -4	0,001	0,001	0,001	0,008		0,001	0,001	0,001	0,001	0,447
1123 0	0,001	0,001	0,001	0,001	0,001		0,900	0,189	0,001	0,001
1123 -1	0,001	0,001	0,001	0,001	0,001	0,900		0,199	0,001	0,001
1123 -2	0,081	0,544	0,017	0,001	0,001	0,189	0,199		0,066	0,001
1123 -3	0,900	0,900	0,900	0,030	0,001	0,001	0,001	0,066		0,001
1123 -4	0,001	0,001	0,001	0,900	0,447	0,001	0,001	0,001	0,001	

signficant ($p \leq 0.05$)

highly significant ($p \leq 0.01$)

A List of Acronyms

AR5 5^{th} Assessment Report

ACE accumulated cyclone energy

ANOVA analysis of variance

AP Average Precision

BIAS bias score

CMIP Coupled Model Intercomparison Project

CNN convolutional neural network

DKRZ Deutsches Klimarechenzentrum

ERA ECMWF Re-Analysis

F false alarm

FAR false alarm rate

GCM global circulation model

H hit

H0 null hypothesis

IBTrACS International Best Track Archive for Climate Stewardship

IPCC Intergovernmental Panel on Climate Change

IQR interquartile range

ITCZ intertropical convergence zone

KDE kernel density estimation

KL Kullback-Leibler

M miss

mAP mean Average Precision

MCS Mesoscale convective system

MDI Maximally Divergent Intervals

NCEI National Centers for Environmental Information

NHC National Hurricane Center

NPP National Polar-orbiting Partnership

OWZ Okubo-Weiss-Zeta

POD probability of detection

PRC precision

PSL pressure at sea level

RH relative humidity

SOS sum of scores

SST sea surface temperature

TC tropical cyclone

TS tropical storm

U-KL unbiased Kullback-Leibler

VIIRS Visible Infrared Imaging Radiometer Suite

Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

München, 10. März 2020

Simon Franz Zitzmann